DOCUMENT RESUME

ED 195 267 IR 009 044

AUTHOR Riggs, Fred W.
TITLE The Design of Glossaries: An Example from "Networking."
PUB DATE Oct 80
NOTE 61p.; Paper presented at the Annual Meeting of the American Society for Information Science (Anaheim, CA, October 1980).

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Design Requirements; *Formative Evaluation; *Glossaries; Guidelines; *Pilot Projects; Social Sciences; *Vocabulary Development
IDENTIFIERS UNESCO

ABSTRACT

Prepared as part of the Unesco INTERCONCEPT pilot project, this study tests the relevance of a set of guidelines for the design of glossaries in the social sciences. The Library of Congress "Glossary of Library Networking" was chosen for this purpose. This paper discusses the analytic problem of glossary design in relation to its intended uses, and explains the annexes which follow: (1) an extract from the networking glossary; (2) the results of a term tracking exercise; (3) a chained sequence of records using materials of the previous annex; (4) a transitional classified arrangement of the same records; (5) a revised classification of the same concept; (6) an index to the terms contained in Annexes B to E; and (7) a final index of the technical terms required in glossary design. (Author/RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195267

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE DESIGN OF GLOSSARIES:
An Example from "Networking"

by

Fred W. Riggs

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Fred W. Riggs

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR009044

DRAFT: Not for Quotation
Please send comments to
the author directly.

THE DESIGN OF GLOSSARIES

An Example from "Networking"

by Fred W. Riggs

CONTENTS:

Introduction 1

PART ONE: SOME PROBLEMS

The Tracking Problem 3
marginals and sleepers 5
the elasticity of definitions 7

Polysemy and Synonymy 10
polysemy as a problem 11
synonymy as a problem 14
semantic modes 18
concluding observations: neologisms 21

PART TWO: TOWARD SOLUTIONS

The Tracking Exercise 27
the inverted entry (or record) 28
chaining records 30

Classifying Records 31
annex D 33
annex E 34

Indexing 37

Conclusion 38

ANNEXES

A: "A Glossary for Library Networking"

B: Tracked Arrangement

C: Chained Arrangement

D: A Transitional Classified Arrangement

E. A Revised Classified Arrangement

F. An Index of Terms in Annexes B to E

G. An Index of Technical Terms needed in Glossary Design

Political Science Department
University of Hawaii
Honolulu, Hawaii 96822
May 1980

ABSTRACT

The INTERCONCEPT Pilot Project, sponsored by UNESCO, has led to the formulation of a set of guidelines for the design of glossaries in the social sciences. A report on this subject, by the author, will be published in UNESCO's series entitled: "Reports and Papers in the Social Sciences." In order to provide a test of the relevance and applicability of these guidelines to an actual glossary project, it was decided to make use of a conveniently available and authoritative glossary as the basis for a preliminary empirical analysis. The "Glossary for Library Networking," published by the Library of Congress, in 1978, on the basis of a document prepared by Dataflow Systems, Inc., was selected for this purpose.

The paper is in two parts, and a set of annexes. Part I sets forth the analytic problem of glossary design in relation to its intended uses, making a major distinction between the requirements of scholars engaged in the production of knowledge (who need univocal and multinomial terms) and the requirements of indexers and information specialists (who need uninomial and multivocal terms). The typical design features selected by subject specialists engaged in the preparation of glossaries are those that have been developed for indexers in the form of a controlled vocabulary or "thesaurus." Their inappropriateness for the design of glossaries intended to serve the needs of writers is explained. Additional information on the problem of "tracking" supports this mode of analysis. Part II explains the annexes which follow.

The annexes contain: A) an extract from the networking glossary, including its introduction and sample entries; B) the results of a tracking exercise based on the definitions of two terms, 'sink,' and 'source' ; C) a chained sequence of records using the material identified in Annex B; D) a transitional classified arrangement of the same records; E) a revised classification of the same concept introducing some new terms for illustrative purposes, and modifying the definitions accordingly; F) an index to the terms contained in Annexes B to E; and G) a final index of the technical terms required in glossary design that were introduced and defined in Part I.

Material contained in this essay supplements the author's report to UNESCO on the INTERCONCEPT project, and provides illustrative material and guidelines for social scientists interested in preparing glossaries in their own fields of specialization. It will be used and discussed at the conference on conceptual and terminological analysis to be held in Bielefeld, Germany, May 25-27, 1980, under COCTA sponsorship.

THE DESIGN OF GLOSSARIES: An Example from "Networking"

by Fred W. Riggs, University of Hawaii
April 1980

INTRODUCTION: A TYPICAL GLOSSARY

0.00 A typical format for glossaries is illustrated in Annex A by means of several pages extracted from a recent and authoritative publication of the Library of Congress. As indicated in the "Introduction" to the glossary, leading authorities working in the field of library networking were consulted. We shall, therefore, take the definitions as authoritative--no effort will be made here to criticize or evaluate them. Rather, attention will be focussed on the principles of glossary design manifested in the text, and the problems they respond to or generate. Scrutiny of the materials given in Annex A will show that the following design principles have been used:

0.01 Terms-to-concepts. The substantive entries consist of a defined term followed by one or more concept definitions. This structure is, essentially, an abbreviated version of the standard format used by lexicographers in their dictionary definitions of words. Supplementary information, such as bibliographic data and notations, is not included in the glossary.

0.02 Synonyms. In a "concept-to-terms" arrangement, each defined concept would be designated by one or more synonymous terms. However, the glossary provides only one term per concept, in almost all cases. In the exceptional cases where synonyms are recognized, one is selected to serve as the entry term, and the synonyms are given with "see" references, instructing the user to go to the preferred entry term. For illustration, see the cross-reference at "state network" in Annex A. This practice reflects standard usage in thesaurus design.

0.03 Alphabetical Order. All entries (both main entries and cross-references) are arranged in alphabetical order, by the defined or referenced terms. This is the characteristic sequence of entries used in dictionaries and most thesauruses.

0.04 Multivocalism. As in the lexical paradigm, it is also accepted that any given word may be multivocal, i.e. that it can have more than one sense--but, of course, only senses relevant to library networking are included. To illustrate, 'source' and 'sink' are ordinary English words with many meanings. Two meanings for each of these terms--meanings not given in general dictionaries--are attributed to them in the glossary. This practice reflects a normal predisposition in the social sciences: assign new meanings to familiar words, but do not coin new words, as natural scientists typically do. If a new word must be coined, avoid New Latin

constructions, but use acronyms, if necessary. The word 'TELPAK' in Annex A gives us an illustration. Other examples found in the glossary are 'WATS' and 'modem.' An exceptional New Latin coinage is 'quadraplanar.' Virtually all other terms given in the glossary use familiar words--i.e. words found in general dictionaries.

0.05 Entailed Terms. All of the words used in the definitions contained in a glossary may be referred to as *entailed terms.* Inspection of Annex A shows that there are two types of entailed terms: those that are defined in the glossary, and others that are not. All the entailed terms that are underscored are defined in their alphabetical order in the glossary. Most of the entailed terms that are not underscored can be understood by reference to their definitions in a general dictionary.

0.06 Defined Terms. All of the terms that are defined in the glossary may be referred to as *defined terms.* Their meanings, for library networkers, are not given in general dictionaries. If they were, it would be redundant to define them in a technical glossary.

0.07 Primitive Terms. All of the terms used in the glossary's definitions that may be understood by means of their definitions, as given in a general dictionary, can be referred to as *primitive terms.* Since the word, 'primitive,' has other meanings, it is important to remember the definition of "primitive terms" given here. Most of the entailed terms given in the glossary that are not underscored are primitive terms. However, not all of the unmarked entailed terms (i.e. those that are not underlined) are primitive terms. Some of the words used in the definitions--e.g. 'data,' 'node,' and 'network,' as we shall see--are not underscored, even though they are defined terms in the networking glossary.

0.08 Related terms. Various relationships between concepts given in the glossary are indicated by "see also" references, given at the end of some definitions. For example, the record on 'source[2]' instructs the user to "See also: sink." The relation of 'sink' and 'source' is apparently one of complementarity. However, in the case of 'text[1]' we are told to "see also: message text". In this instance, the two terms are apparently used as synonyms, as near-synonyms, or perhaps in a genus-species relation.

0.09 Compound terms. Many of the defined terms given in the glossary are made up of two or more words, such as 'site data,' 'star network,' 'union catalog,' and 'two-way simultaneous operation.' In biology such terms may be called 'binomials' and 'polynomials,' depending on their use of two, or three or more, words. Here we may more simply call them *compound terms.* When a compound term is used as an entailed term in the networking glossary, it can be identified as such if it occurs

* Starred and underlined expressions are new terms: their definitions are given in the text and also in Annex F.

on a single line of the text because it is underlined without any break between its component words. However, when compound terms are distributed between two lines, it is not easy to tell by simple inspection whether it is a single compound term, or two defined terms in sequence, forming a phrase. For example, the terms 'message,' 'text,' and 'source' are defined terms, and so is 'message text,' but not 'message' 'source.' Therefore, if "message source" is encountered in the glossary, it must be read as a phrase composed of two terms, not as a single compound term. This may seem like a very fine point until one reflects on the meaning of a compound term that is quite different from that of a phrase composed of the same words: an example is 'full stop', which does not mean 'full' 'stop.' Put differently, if the meaning of 'message text' is the same as the meaning of 'message' 'text,' then it is redundant to define it in a separate entry.

Having now described the networking glossary's design features, we may turn to an analysis of some of the problems that it poses.

PART ONE: SOME PROBLEMS

1.00 Two major, interrelated sets of problems in glossary design are well illustrated by the materials contained in the networking glossary, as shown in Annex A. The first set of problems to be examined concerns the sequence of records which makes up a glossary. The unintended consequences of an alphabetical arrangements will be explored. A second set of problems concerns the sequence of elements found in each glossary record. The unintended consequences of the term-to-definitions sequence will be examined. We shall start by looking at the problems generated by an alphabetical arrangement of records, as reflected in the problem of "tracking."

1. The Tracking Problem

1.10 Independence. In order to explain the problem of "tracking" we must first distinguish between independent* and interdependent* definitions. If one knows all of the terms used in a definition, then it can be understood independently*of other definitions contained in a dictionary. Since lexicographers report, in general dictionaries, on the established meanings of words in current use, and only use such words in their definitions, it follows that dictionary (or lexical) definitions are independent.

1.11 Interdependent Definitions. By contrast, in glossaries where the purpose served is to introduce users to a set of new concepts and their terms, definitions are typically interdependent.* This occurs because some of the entailed words used in definitions are defined words, and hence it is necessary, quite often, to look them up before the meaning of an entry can be grasped. In order to explain the reasons for interdependent definitions more precisely, we need to distinguish between

primitive* and derivative* terms. In lexical usage, a "primitive" is a radical or root word, and "derivatives" are words formed, by derivation, from such primitives. This concept is structural, and of course lexical derivatives, as well as lexical primitives, are found in general dictionaries.

1.12 Semantic Derivatives. By contrast, we need a different, though related, concept that is semantic rather than lexical. We have already defined, #0.07, "primitive terms" as words with their meanings, as given in a general dictionary. A new semantic sense of derivative* is that of a word with a sense not given in a general dictionary. Since familiar words may be used for new senses, clearly the word-form used as a semantic derivative may be found in general dictionaries, but not the new sense for which it is used. To show that a word is a derivative, therefore, one needs to find that it is used for a sense not included in its general dictionary definition. It is important to remember that whereas lexical derivatives (word-forms) may be found in general dictionaries, by definition semantic derivatives will not be found in such dictionaries. It will be seen that the compound term, 'semantic derivative' is itself a semantic derivative. In general usage, the terms 'primitive' and 'derivative' refer to the lexical senses, i.e. for the word-forms employed. By contrast, in this essay, 'primitive' and 'derivative' will be used only in their new semantic senses, to distinguish between terms whose definitions are given in general dictionaries, and those whose definitions are not found in such dictionaries. (To operationalize the concept, a particular general dictionary has to be used: here Webster's New International Dictionary will be the authority)

1.13 Derivatives and Interdependence. We can now state the defining characteristic of interdependent definitions* more formally: any definition that contains one or more derivative term is interdependent. The point, of course, is that in order to understand the meaning of interdependent definitions, one must find, outside a general dictionary, the meaning of whatever derivative terms are used as entailed terms in the definitions.

1.14 Tracking. We can now define tracking* as the procedure involved in hunting down the meaning of derivative terms and reducing them to primitive terms, i.e. by discovering the ordinary language terms that explicate their meanings. Perhaps the best way to clarify this concept is to offer an illustration. The word, 'data,' for example, is defined in the networking glossary as:

(1) a representation to which meaning can be assigned.

1.15 Primary Derivative. This sense of 'data' is not found in Webster's general dictionary, and we have to conclude that it is used, in the glossary, as a derivative, not as a primary term. However, it is a primary derivative* in the sense that it can be defined by the use of primitives only.

1.16 Secondary Derivative. Any definition that contains a primary derivative may be said to establish the meaning of a secondary derivative*--provided it does not also contain any secondary or higher level derivatives. To illustrate, consider the definition of one of the two senses of 'channel' given in the networking glossary, as follows:

(2) a means of one-way transmission of data.

The word, 'channel,' when used in this sense, is therefore a secondary derivative.

1.17 Higher Level Derivatives. If we track the term 'sink' as defined in the networking glossary, we shall find that it is a higher level derivative* because one has to go through several (actually seven) steps of tracking in order to reduce all the derivatives found in its two definitions to primitives. This illustrates the general problem of tracking which is inherent in the design of any glossary for new concepts.

1.18 Alphabetization. Although it is always an effort to track the meanings of derivative terms used in interdependent definitions, this effort is magnified when the records giving the definitions in a glossary are arranged in alphabetical order. Instead of bringing the derivative terms close together in a logical sequence that would facilitate the process of learning them, they are scattered according to the accidental orthography of the key words that are used as entry terms.

Marginals and Sleepers

1.20 Not all of the derivative terms used in a glossary are defined in it, a fact that further complicates the task of tracking. The reasons become apparent when we note that glossaries are prepared in an intellectual context of discovery within which many new concepts have already become well established in related subject fields. Accordingly, glossologists may assume that their audience will be familiar with these derivative terms, and they will not bother to define them. In other words, not all the derivative terms in a glossary are defined terms--at least, not necessarily. Let us speak of any derivative terms that are treated as part of the necessary and new vocabulary (nomenclature) of a subject field as its core derivatives*. By contrast, all derivatives used in a given subject field that are not core derivatives may be referred to as its marginal derivatives*. It would, no doubt, be very convenient if we could assume that all the derivatives used as entailed terms in a glossary were core derivatives, and that they were marked by

some means, like underscoring, and were defined in the glossary. In fact, however, many of the entailed terms used in glossaries are likely to be marginal derivatives, whose definitions are not given, nor may they be found in general dictionaries.

1.21 Marking Marginals. Although I can not determine the extent to which definitions contained in the networking glossary use undefined marginal terms, there is a simple procedure which can be used to identify and define marginal derivatives. It simply requires that all derivative terms used in a glossary's definitions be identified by its authors and suitably marked--e.g. by underlining, italicization, or some other suitable means. Among these derivatives, some will be core terms, and they should be defined in the main body of the glossary. Other derivatives (the marginals) need not be defined in the main body, but information should be provided to help users who are not familiar with them to track down their meanings. Such tracking information may take the form of a reference number guiding users to another glossary or specialized dictionary in which the marginal terms are defined. Another procedure involves the addition of definitions to the index, following the index terms which guide users to any places where the term in question has been used.

1.22 Tracking Unmarked Marginals. Since marginal terms are not marked in the networking glossary, however, it is impossible for users to identify them without making a dictionary search first in order to determine whether or not the probable sense of the term, in context, corresponds to one of the dictionary senses of the word. If the marginal terms were unfamiliar words, it would be easy enough to spot them. However, in the networking glossary virtually all of the terms used for new concepts are familiar words, with the result that it may be difficult to track down the meanings of those that are used as derivatives, but are unmarked.

1.23 An Example. By way of illustration, consider the status of 'representation,' a word used in the definition of 'data,' as given in (1). A careful reading of the definitions offered for ten senses of 'representation' in Webster's dictionary failed to uncover one that would appear to fit precisely the meaning of this word as it appears in (1), yet the word is not marked nor defined in the glossary. Although readers may well infer its meanings from the context of its use, it is also quite possible to misconstrue its meaning which seems not to conform to any of its many general dictionary senses.

1.24 To recapitulate, the problem of tracking poses a double hazard for glossary users. First, it may be necessary to track the meanings of the defined terms that are clearly marked, and this is difficult enough. In addition, however, it may be important to track the meanings of unmarked marginal terms, a task that can be extremely arduous.

1.25 Sleepers. In addition to the unmarked marginal terms found in glossary definitions, we may sometimes encounter unmarked core terms. Such terms are, in principle, defined in the glossary so that it is much easier to track them than it is to find the meanings of marginal derivatives. To illustrate, the word 'data' used in the definition of 'channel' as given in the networking glossary is not underlined--the underlining was added in (2). However, as we have seen in (1), 'data' is a defined term in the networking glossary. In order to designate unmarked defined terms--and they are among the core derivatives of a subject field--we may refer to them as sleepers*. Clearly it is easier to find the sleepers in a glossary than it is to discover the unmarked marginal terms. Nevertheless, the failure of glossographers to mark all the defined terms that are used in their definitions adds to the difficulties facing users who might even suppose that such words were being used as primitives, rather than as derivatives.

The Elasticity of Definitions

1.30 Technical Jargon vs. Ordinary Language. Some critics of the use of technical terms seem to think that the tracking problem can be overcome by the systematic substitution of semantic primitives for derivative terms in specialized contexts. An example of this opinion can be found in the following statement, taken from a report by UNESCO's General Information Programme, that was prepared for the UNISIST II conference in Paris, May/June 1979.

> The increasing specialization of scientific research has led to the development of technical jargon which hampers understanding even among specialists when they have to face up to multidisciplinary problems. For communication to many audiences STI has to be converted into ordinary language. (#109, p.99, in PGI/UNISIST II/4)

1.31 Term Replacement. In general, whenever one term replaces another in a definition, we may speak of term replacement*. Replacing technical jargon with ordinary language would be an example of term replacement. Whenever a defined term is used as an entailed term, it is, in principle, possible to replace it with that term's definition. By this means, the derivational level of a definition can be lowered. Consider, for example, the following set of definitions, starting with (1) and (2) as given above:

(3) primary derivation: data: a representation to which meaning can be assigned

(4) secondary derivation: channel^2: a means of one-way transmission of data

(5) 3rd level derivation: source^1: a computer...that receives and processes data from a connected channel

(6) 4th level derivation: channel^1: the connection between a message source and a message sink

1.32 Condensation and Stretching. Each of these definitions is condensed* insofar as one of its defining characteristics is represented by a derivative term, but it can be stretched* if that term is replaced by its definition, as follows:

(7) 1-step reduction: channel^2: a means of one-way transmission of representations to which meaning can be assigned

Since the definition of 'channel^2 given in (4) is a secondary derivative, it can be reduced to a primary derivative in one step, as shown in (7). By this means, the definition of this term is given in ordinary language, and the technical term, 'data,' previously used in the definition has been eliminated.

1.33 A 2-step Reduction. The one-step reduction of a secondary derivative is not difficult to perform, but the two-step reduction of a third-level derivative that is required to define it in ordinary language is more complicated. Consider, for example, the term replacement required to reduce the definition of 'source^1', as given in (5):

(8) 2-step reduction: source^1: a computer...that receives and processes representations to which meaning can be assigned, that have been received from a means of transmission of representations to which to which meanings can be assigned

1.34 A 3-step Reduction. By the same procedure, we could reduce the definition of 'channel^1' to ordinary language by replacing, in its definition (6), all of the text of (8) for the word 'source,'--plus an equally long stretched out definition to replace 'sink.' Notice also that the definition of 'channel^2' (7), in this reduction becomes a component of the definition of 'channel^1.' The possibilities of confusion involved in this kind of a reduction whereby "ordinary language" replaces "technical jargon" boggle the imagination. No attempt will be made to produce such a 3-step reduction--but the reader may make an attempt to do it!

In the previous paragraphs, the term reduction* has been used as a derivative, to refer to the process of transforming a definition by stretching it, thereby lowering the level of derivation, moving from technical terms to ordinary language. Thus the process of derivation is the opposite of reduction.

1.35 Trade-offs. Even a two-step reduction (stretching) as (8) shows, leads to long and complex definitions which may well be quite difficult to understand. The definition of 'source' given in (5) appears to be much clearer than the definition of the same term given in (8)--but it presupposes an understanding of the meaning of two defined terms and requires a two-step derivation. We may conclude that although the definitions given in an alphabetized glossary can be made independent by stretching, this may generate unintelligible definitions. Indeed, one of the reasons why so many of the definitions offered within the body of subject field essays are so incomprehensibly complex is because they have been unmercifully

stretched in order to reduce them to primitive terms.

1.36 Intelligibility can be greatly enhanced by the condensation of definitions, replacing primitive terms with derivatives. However, the more this is done, the more serious become the tracking problems. Accordingly it is important to design glossaries so that the problems involved in tracking entailed terms can be minimized. We shall see, I believe, that these problems can be solved much more efficiently by arranging glossary records in a systematic order than by attempting to reduce them through stretching.

1.37 Elasticity. If we think of the elasticity* of definitions as a measure of the degree to which they can be stretched so as to reduce the cost of tracking their entailed terms, we can formulate some interesting hypotheses that are relevant to the design of glossaries. In other words, the more elastic a definition, the easier it is to reduce its level of derivation without making it more difficult to understand. Accordingly, the more elastic the definitions given in a glossary, the easier it is to convert their semantic derivatives into primitives--i.e. to replace technical jargon with ordinary language. This can be regarded as the ideally preferable means to solve the tracking problem.

1.38 Inelasticity. Quite often, however--as in the networking glossary--we find that the definitions given are quite inelastic*, i.e. that efforts to stretch them so as to reduce their derivativeness make them, actually, less rather than more intelligible--as one may see by comparing (5) with (8), or (6) with the imagined definition proposed in #1.34. Put differently, the more elastic the definitions found in a glossary, the easier it is to solve the tracking problem by stretching them so as to reduce their level of derivativeness. Conversely, the more inelastic the definitions found in a glossary, the more they need to be condensed, thereby increasing their level of derivation and making it necessary to find other means to solve the tracking problem.

1.39 Classification vs. Alphabetization. These considerations lead us to state that, if other means of solving the tracking problem can be found, then the increased clarity of definitions that can be secured by condensation will be fully justified. What alternatives to reduction are available? As I shall try to demonstrate, the substitution of systematically ordered records for an alphabetical arrangement can reduce tracking costs, thereby providing a substitute for the stretching of definitions. In order to discuss the classification of glossary records, however, we need to consider their internal structure. To understand this structure, however, we must also consider the interrelated problems of polysemy and synonymy, and how these problems affect the design of individual glossary records. Looking at these problems, we shall also see that solutions

which are designed to meet the needs of information retrieval are not well suited to the requirements of scholars engaged in the production of knowledge.

2. Polysemy and Synonymy

2.00 The interrelatedness of polysemy and synonymy as they affect the design of glossary records and thesauruses is rarely noticed. Some people see no problem here. They accept as non-problemmatical the fact that one word can have a multiplicity of meanings, and that one concept can be designated by a variety of terms. By contrast, there are utopians who treat both polysemy and synonymy as unrelated problems, each of which can be solved independently in the name of "univocalism," interpreted to mean one concept per term and also one term per concept. In my opinion the interrelatedness of these two phenomena leads to a situation in which one can have terms that designate just one concept only if one accepts several terms per concept; and reciprocally, the price of having only one term per concept is the necessity of accepting terms that designate more than one concept. The reasons for this trade-off will be explained below, and their implications for the design of glossaries will be stressed.

2.01 The Lexical Orientation. The orientation which accepts both polysemy and synonymy as non-problematical is lexical. Lexicographers devote much attention to the identification of all the established senses of each word presented in a general dictionary and, conversely, they recognize synonymies and seek to distinguish the marginally differentiated connotations of each synonym in every set.

2.02 Information Specialists. By contrast, indexers need to have a single term for each concept, since otherwise anyone retrieving material on a given subject would have to work with synonymies, searching under several synonyms. In order to avoid synonyms, however, a price must be paid in the form of polysemy, for reasons to be explained below. This trade-off is reflected in the design of thesauruses, the tool par excellence for indexing purposes.

2.03 Knowledge Production. Scholars engaged in the production of knowledge, by contrast, need glossaries which reverse the priorities appropriate to information retrieval. When one is writing in a technical or scientific mode, it is important to be able to express concepts unambiguously, and for this reason scholars search for terms that can be used to express, in context, one and only one concept. They have no objection to synonyms, however--indeed, once the identity of an intended concept has been clearly indicated, it is a relief to be able to use a variety of synonyms to designate it. Only scholarly writers, of course, have this need. Literary, political, and many other forms of rhetoric often choose, deliberately, to use ambiguous metaphors and overloaded terms.

2.04 "Glossology." Since the subject specialists who prepare glossaries are not, themselves, specialists in "glossology," it is scarcely surprising that they use models that were designed to meet needs different from their own. The fact that glossology*--defined as the theory and knowledge required to prepare glossaries--is itself not a primitive term may be taken as indicative of the lack of expertise in this promising field of specialization. In the absence of glossology, therefore, glossographers--those engaged in the preparation of glossaries--draw on the models supplied by lexicographers (for the compilation of dictionaries) and by information scientists (for the creation of thesauri). Why these models are inappropriate will become more apparent when we pay attention, in detail, to the problems of polysemy and synonymy.

Polysemy as a Problem

2.10 Polysemy should first be distinguished from homonymy. Lexicologists make a sharp distinction, which is reflected in any general dictionary. If two different words possess the same form, they are recorded in two separate dictionary entries. According to Webster's dictionary, for example, there are two homonyms which have the form, 'sink.' One is a verb, which may be declined as 'sink, sank, sunk.' By contrast, there is another word, 'sink,' which is a noun taking the forms, 'a sink, 2 sinks'. Each of these homonyms, however, is also a polyseme possessing a variety of senses. The verb, 'sink,' for example, is reported by Webster's dictionary to have 21 main senses, while the noun, 'sink,' has ten major senses. In all of these senses, of course, 'sink' is a polysemantic primitive, by definition. Note that the term, 'polyseme,' is used to designate a word (which may or may not also be a homonym) that has a variety of senses in all possible fields of knowledge. In any given field of knowledge, however, the same word may have only one meaning. Thus, paradoxically, any polyseme can be univocal in a particular subject field. The word 'station,' for example, is lexically a polyseme, but in library networking, according to the glossary, it designates only one concept, and is therefore univocal.

2.11 Polysemantic Derivatives. Any word may acquire more senses than those given to it in ordinary usage, as reported in a general dictionary. Such terms are semantic derivatives, as explained above in #1.12. Semantic derivatives may, of course be polysemantic within the subject field concerned--or they may be univocal. The term, 'sink ,' by contrast with 'station,' is used for two concepts in the field of library networking, according to the networking glossary. Since neither of these senses of 'sink' is reported in a general dictionary, we must treat them as derivatives. And since the word can designate two different concepts within the field of networking, it is a polysemantic derivative.

2.12 Multivocalism and Polysemy. The words 'multivocalism' and 'polysemy' can be treated as synonyms, expressing the idea of a single word having several meanings. In the context of our analysis of glossary design, however, it is useful to make a conceptual distinction which can be differentiated by the use of both words. In the ordinary sense, a word is polysemantic if it has different meanings in any and all contexts of usage. For our purposes, however, what is important is whether or not a word designates more than one concept with a given context of use, e.g. the subject field covered by a glossary. We have, accordingly, a general concept of the meanings of words in all possible contexts, and their meanings in a special context. I shall restrict the meaning, in this paper, of multivocalism* to the sense of being polysemantic in a given subject field, e.g. "library networking." The word 'polysemy' can then be used in its ordinary sense, as a primitive term, according to its general dictionary definition. See again #0.04 above.

2.13 Univocalism and Monosemy. In #2.00 the word 'univocalism' was used in a popularized sense to mean not only that a term would have one meaning, but also that the designated concept would have only one term. This usage, however, is rather loose, and we need to tighten up our nomenclature to handle the glossary problems encountered in the networking glossary. Actually, we have two words to consider, and they may be regarded as antonyms for the terms defined in #2.12: namely 'univocalism' and 'monosemy.' They can, of course, be treated as synonyms, both designating the concept of a term with one meaning only. However, we need to distinguish between the possibility that a term will have one meaning only in a given context of usage, and the probability that it will have more than one meaning in all contexts of use. In the unlikely case of a term that has only one possible meaning in all contexts, let us refer to this situation as one of monosemy*, by contrast with polysemy--no doubt some New Latin terms in biology have this property. However, such terms are also univocal* in the narrower sense to be defined here, namely that of terms that have only one meaning within a given context of usage, such as a scholarly subject field. Clearly words like'station', which are by no means monosemantic, can be univocal in a narrower field, like networking. If we can sustain this distinction, it will be easy to remember, since the pair of terms, 'polysemy/monosemy' refer alike to all contexts of the use of a word; whereas 'multivocal/univocal'--as defined here--would refer to the meanings of words as they occur in a specified context of usage, such as a subject field.

2.14 Univocal and Multivocal Terms. To be more precise, we can add the word, 'term' so as to create designations for a word that is used to signify one or more than one concept within a specified subject field. To illustrate, as we have seen, 'station' is a univocal term* in the context of networking, and 'sink' is a

multivocal term*, at least according to the evidence provided by the networking glossary. Both words are, of course, polysemes--but we can now see that this makes no real difference when it comes to the problems of glossology. We may therefore conclude that the so-called problem of "polysemy" can be reduced to the narrower and more manageable problem of "multivocalism."

2.15 Different Orientations. We can now see more clearly how the problem of "polysemy" is viewed from differing points of view. Lexicologists do not make the distinction drawn here between "polysemy" and "multivocalism" for quite good reasons--they undertake to investigate and report all the senses of a word as it may be used in every subject field. Accordingly their interest coincides with the general phenomenon of polysemy. Moreover, they do not see it as a problem, but rather as a challenge to discover and identify and define the various meanings of each word, treated as a polyseme. In the preparation of subject field glossaries, by contrast, it is important, to recognize terms that are multivocal--but the fact that they are polysemes can be ignored. This distinction which is opaque to lexicology now appears to be important to glossology. For information scientists, by contrast, the fact that terms may be multivocal is certainly a matter of concern. However, as we shall see, the problems of synonymy take priority, and if the solution to this problem requires that multivocalism be ignored, then this becomes accepted practice. The results can be seen in the networking glossary, where many words are presented as multivocal terms. How the problem of synonymy has led to this outcome will be discussed below.

2.16 The Goal of Unambiguous Writing. For producers of scholarly knowledge, however, multivocalism generates a major problem. If one wishes to write about a given subject unambiguously, but the only available terms convey unintended meanings within the subject field at hand, then scholars find themselves trapped by a difficulty. In order to make their intentions clear, they often find themselves spending an inordinate amount of time and effort defining terms and explaining what they mean. In the process, they frequently add new meanings to familiar words, thus increasing the burden of term overloading and of multivocalism. If glossaries were designed to solve the problem of multivocalism that confronts scholars engaged in the production of knowledge, rather than to solve the different though related problem of synonymy, which is of primary importance to indexers, scholars would find their difficulties alleviated--and they would be much less tempted to contribute to the proliferation of multivocalism.

2.17 The Choice of a Model. If it is true, as I suspect, that most glossaries obey the same design principles as those manifested in the networking glossary, then we have to ask why this is so. Why should glossologists not select a model

more suited to the requirements of scholars engaged in writing documents than to those of indexers working on written materials? It is only a surface explanation to say that, in the absence of a theory of glossology, glossographers have naturally turned to the design models provided by lexicographers, and by the creators of thesauri. A deeper level of explanation will become apparent if we understand why the information scientists and linguists who have, with great care, designed the models emulated by glossographers, did indeed establish the criteria that they chose. I believe the basis for choosing these criteria will become apparent if we learn more about the problem of synonymy, and its solution.

Synonymy as a Problem

2.20 Two words are said to be synonyms if their meanings are similar. The word, 'quasi-synonym,' is sometimes used to suggest that if two words are true synonyms, their meanings must be identical. Lexically speaking, however, in view of the fact that almost all words are polysemantic, it is virtually impossible to find two words all of whose senses are identical, taking into account their value connotations as well as their objective denotations. Because of this, lexicographers often present sets of synonyms--which they call "synonymies"--so as to distinguish between the differing meanings of each synonym in the set. To illustrate, Webster's dictionary entry for the word, 'prevent,' includes, following the abbreviation, 'syn.' the words: 'preclude, obviate, avert, and ward(off)...' In the text that follows, the distinctive meanings of each of these synonyms is carefully distinguished from the comparable meanings of the other words in this synonymy. Glossographers typically reject this feature of lexicographic practice, however, preferring the approach taken by information specialists in the preparation of thesauruses.

2.21 Retrieval. For indexing purposes, synonyms pose a much greater problem than polysemy. In order to index a given text, it is necessary to choose a single term for each subject or concept indexed. If several members of a synonymy were to be used to index a single topic, the user would have to identify and then search under all the synonyms for a given concept in order to locate the relevant literature. To meet this difficulty, information specialists have established controlled vocabularies that take the form of a thesaurus. In such a vocabulary, only one member of each synonymy is selected, and cross references are used to direct the user from any of the other synonyms to the one selected. If, for example, one could use "data," "information," and "message" as synonyms, then a thesaurus might select one of these words, e.g. "data," and after each of the others place a cross-reference that would read: "SEE: data." As a result, of course, any controlled vocabulary involves the selection of a preferred term for each concept to be indexed, so that for each indexed subject there will be one and only one index term.

If we look at the structure of the networking glossary -- see Annex A, and #0.02 above--we shall readily see that this is the paradigm that has been followed.

2.22 Writing. As has already been pointed out in #2.03, scholars engaged in the production of knowledge welcome available synonyms that also permit them to express their intentions unambiguously. The kind of tool that they need to facilitate their work, by contrast with the needs of indexers, is therefore a glossary that would supply a list of synonymous terms that can be used unambiguously to designate whatever concepts they wish to use. Clearly this is a different preference, not to say priority, from the one chosen by indexers and information specialists as explained above.

2.23 Uninomial and Multinomial Concepts. The contrasting requirements of indexing and writing can be discussed more easily if we introduce two new technical terms, as follows: 'uninomial' and 'multinomial.' The term uninomial* is actually a primitive, since Webster's unabridged dictionary defines it as "a system of nomenclature in which items have single names." However, its use is so uncommon that it may be treated here as a new term, if not a derivative. More importantly, it is indeed an innovation to use the term, multinomial* as the antonym of 'uninomial,' i.e. to represent a system of nomenclature in which items have more than one name--to paraphrase Webster's dictionary. The word, 'multinomial,' does indeed appear in the dictionary, but as a synonym for 'polynomial.' However, as we have already seen in #0.09, the terms 'binomial' and 'polynomial' are used, as primitives, to designate terms that contain two, or more than two, words. Here we need a different concept, namely concepts with two or more terms. It seems appropriate to use 'multinomial' for this purpose, and it should not be difficult to remember the special meaning given the word in this context.

2.24 Synonymous Terms. We must now distinguish rather sharply between"synonyms" and "synonymous terms." The former term is used lexically to refer to properties of words as they are found in a language. The latter term, by contrast, is used glossographically, i.e. to refer to sets of terms that can be used to designate any single concept. Although a multinomial concept has, by definition, several terms, these terms need not be synonyms, since each term in the set can signify various other meanings, due to its polysemantic properties. Put differently, if a term is univocal, then only one of its possible meanings will be intended when it is used in the context of a given subject field, even though the word, in other contexts, has various other possible meanings. Consequently two different terms for a single concept may be synonymous terms, but not synonyms. To illustrate, the words 'computer' and 'terminal' are clearly not synonyms, as one may readily determine by looking up their definitions in a general dictionary. If, however,

in the context of networking they could be used to designate the same referent, as they are not, then they might be treated as synonymous terms. Since all of the concepts defined in the networking glossary are uninomial, we cannot provide any real examples of synonymous terms from this source, but for illustrative purposes a variety of synonymous terms have been invented, and they are entered in Annex E, as we shall see below.

2.25 Term Family. If in a given context of usage, therefore, several different terms can be used unambiguously, even if they are not synonyms, to designate a single concept, we may refer to this set of synonymous terms as a term family*. By including members of a term family in each glossary record, users will learn that they have a choice among these terms whenever they wish to use the concept that they designate. In order to make such choices, however, users will also want to know something about the circumstances under which it is appropriate to use each member of the family.

2.26 Members of a Term Family. The simplest distinction that we can make between different members of a single term family involves frequency of use by specialists in the subject field of a glossary. Obviously some words are much more familiar than others. When neologisms are introduced to a field, by stipulation, completely new terms may be proposed that users may, initially, tend to forget--or even reject as unacceptable. (An example will be offered in #2.27) In order to be able to talk about this distinction more clearly, let us use familiar term* in a special sense to designate any member of a term family that is relatively well known to specialists in the subject field community addressed by a glossary; and we shall use unfamiliar term* to designate the contrary concept. Obviously there is no simple dichotomy between familiar and unfamiliar terms--rather, these are polar concepts, with varying intermediate degrees of greater or less familiarity among the members of a term family. Accordingly, it may be necessary to characterize any given term as relatively more or less familiar or unfamiliar. Obviously the more familiar a term, the easier it will be for users to remember it, and the more unfamiliar, the more effort will be required--at first--to learn it. It is important, also, to remember that we refer here, not to the familiarity/unfamiliarity of words in all general contexts of their use within a language community. Rather, we refer to the familiarity/unfamiliarity of a given term as it is used by specialists in a given discourse community to refer to the special concepts of their own subject field.

2.27 The Format of a Glossary Record. The notion of a term family suggests a new way of looking at the contents of a glossary record. In the usual format, as

illustrated in Annex A, we see that each record contains two parts, as noted in #0.01: namely a single term followed by one or more concept "definitions." To be more precise, the term to be defined is sometimes called the "definiendum," and each of the defining texts that follows is called a "definiens." The record format, then, consists of a definiendum followed by one or more definientia that define the senses of the word as it is used in the field of library networking. This rhetoric makes it seem that the thing to be defined is a word, and that which defines the word is a text. No doubt this rhetoric precisely fits the requirements of lexicographers when they prepare dictionaries. However, it falls short of meeting the needs of glossographers. Consider the paradox that occurs if we attach a cluster of terms to a record, in the form of a term family. Would we then have, instead of a single definiendum, a set of definienda? But in the usual treatments of definitions, there can be only one definiendum. The way out of this apparent paradox is simply to recognize that what is to be defined is not a word but a concept. Since difficulties sometimes arise when we try to pin down the meaning of a "concept" by its referent and defining characteristics, we may short circuit this philosophical issue by referring to a defining text* as that portion of a glossary record that corresponds to the definiens of a dictionary entry. However, the function of a defining text in a glossary record differs from the function of a definiens in a dictionary entry. In the latter case, it pins down one sense of a word, but in the former case, it identifies a concept by reference to some defining characteristics--or by identifying its referent. Once we discover a concept by means of its defining text, we may then ask what terms can be used--or are used--to designate this concept. For convenience, let us refer to the set of terms, or the term family, that can be used to designate the meanings set forth in a defining text as a glossonym*. It can be anticipated that some readers will promptly reject 'glossonym' as a formidable neologism--and it certainly is an unfamiliar term. However, any readers who find it difficult to remember or accept this word are invited to suggest synonymous terms for it, thereby creating a new term family for a new concept. The substantive point to remember is that a glossonym is not a definiendum, because it is not a word to be defined, but rather it contains a set of synonymous terms that can be used to designate the meaning of a single defining text. In order to emphasize the difference between the relation of definiendum to definiens in lexical definitions and the contents of a glossary record, it is helpful to reverse the sequence, i.e. put the defining text first, and let the glossonym follow. Of course, it would be possible to let the glossonym of a record precede the defining text, but when this is done, the temptation is very strong to think in terms of the inappropriate lexical paradigm, i.e. of a definiendum followed by a definiens.

2.28 STIC's and STOC's. We can now make a further distinction between two types of members of term families, permitting the division of glossonyms into two parts. The distinction to be made is related to but significantly different from the "familiar/unfamiliar" coding discussed above in #2.26. The essential point is that some terms in a term family may be univocal, whereas others are multivocal. (see #2.14) In order to use a multivocal term unambiguously, it is necessary to supply contextual clues that tell the reader which of several possible meanings of a term is intended. Accordingly, we refer to the multivocal members of a term family as synonymous terms that are unambiguous only when used in context, or STIC* for short. A short acronym is appropriate because it is useful to have a label that can be inserted in glossonyms to indicate that caution must be exercised in order to avoid using the marked terms ambiguously. Since STIC resembles the familiar term, KWIC (for "key word in context") it should be easy to remember. A contrasting type of synonymous term is univocal, and may therefore be used unambiguously out of context, so we call it a STOC*. However, it must be emphasized that univocal terms are not necessarily monosemantic (see #2.13)--i.e., they may have additional meanings outside the subject field of a glossary, but they occur in only one glossonym within the glossary of that field.

2.29 Familiar STIC's and Unfamiliar STOC's. It is surely not always true that a STOC is unfamiliar, but frequently this is the case. By contrast, an important reason for using a STIC is definitely its familiarity. To illustrate, consider the example of 'text' as it is defined in the networking glossary, where it is used to designate three different concepts. In other words, it can be used unambiguously only in contexts that indicate which of the three possible meanings of 'text'--as it is used in library networking--is intended. This task can be simplified by adding a STOC to the glossary record for 'text.' For example, one of the meanings of 'text' might also be termed, 'whole text,' another meaning, 'message text,' and the third meaning, 'running text.' (No doubt specialists in networking would select different STOC's, but the point to be stressed here is that it is quite feasible to select synonymous univocal terms, even if they are not familiar, that can be used in a given subject field for one and only one concept) If such STOC's were to be added to the STIC's for a given defining text, then users would have this choice: "use the familiar multivocal term, 'text,' whenever no ambiguity is likely, but if there is any chance that a statement might be ambiguous, then use the more unfamiliar term, such as 'message text.'"

Semantic Modes

2.30 In order to synthesize our analysis of the relation between polysemy and synonymy--or more precisely between multivocalism and multinomialism--consider

the matrix shown in Figure 1:

FIGURE 1: Semantic Modes Defined

TERMS PER CONCEPT	CONCEPTS PER TERM (in a subject field)	univocal	multivocal
	uninomial	A	B
	multinomial	C	D

2.31 The Dimensions. The horizontal dimension distinguishes between usages in which, within a glossary's subject field, each term designates only one concept (univocal) and those in which it signifies more than one defined concept (multivocal). On the vertical dimension, we distinguish between glossaries in which each concept has only one term (uninomial) and others where it has more than one term (multinomial). The cells in the resulting matrix are, provisionally, designated by the letters, A to D. Something can now be said about each of them.

2.32 The (A) and (D) Cells. The first (A) cell is really empty. Although some idealists advocate the possibility of a nomenclature that is both univocal and uninomial, in my opinion this is both a utopian and quite unnecessary ideal. Although the term, 'univocal,' is sometimes used in this sense, I believe it would be much better to employ a different term, e.g. unisemantic*. As for the fourth cell, (D), it represents the lexicographical norm and practice, namely the definition of all the senses of a word (multivocalism) and the presentation of all synonyms for any given concept (multinomialism). If this approach were to be used in the design of glossaries, it might well be called multisemantic*. However, in my opinion, the multisemantic mode, while quite feasible, is neither helpful to information specialists intent on indexing and retrieving written materials, nor is it useful as a guideto scholars engaged in the production of knowledge in technical subject fields. In short, we may disregard both the unisemantic and the multisemantic modes as guides for the design of glossaries.

2.33 The (B) Cell: Information Needs. Type (B) represents a uninomial but multivocal semantic mode. It provides an appropriate paradigm for information retrieval, and it is well reflected in the design of thesauri. In such controlled vocabularies one term only (uninomial) is selected for each concept--and normally this has to be the most familiar term--even though the word chosen may have (as it often does) a variety of other meanings (multinomial). This semantic mode can well be termed the thesaurus semantic mode* : it is not only appropriate for the design of thesauruses, but it is in fact the basic paradigm for thesaurus construction. By contrast, this paradigm is not suitable as a basis for the design of glossaries that are intended to help scholars write texts. However,

the thesaurus semantic mode is, in fact, often used as a basis for the design of glossaries, as we may illustrate by Annex A, as described above in #0.02 and #0.04.

2.34 The (C) Cell: Scholarly Needs. The needs of scholars, as writers, are much better served by an alternative paradigm (C) that is univocal but multinomial. In other words, for each defined concept in a glossary, one may find several synonymous terms, at least one of which has no other meaning in the context of the glossary's subject field--though it may have other meanings outside that field. For convenience, we may call this paradigm glossary semantic*. It will be illustrated in Annex E--to be described below, in Part II. Unfortunately the glossary semantic mode is rarely used in the actual design of glossaries simply, I believe, because it is not understood, and the thesaurus semantic mode is uncritically accepted by glossographers who are, after all, specialists in the subject field of the glossary, not in glossary design.

2.35 Recapitulation. To summarize our discussion of the four different semantic modes, we can now replace the letters in Figure 1 with the terms offered above, as follows:

FIGURE 2: Semantic Modes Designated

		CONCEPTS PER TERM (in a subject field)	
		univocal	multivocal
TERMS PER CONCEPT	uninomial	unisemantic	thesaurus-semantic
	multinomial	glossary-semantic	multisemantic

2.36 The Thesaurus Semantic. The justification for the thesaurus semantic is well known: uninomialism permits the use of a preferred term for indexing completed texts. If different synonyms were to be used to index a single subject or concept, users would be unnecessarily confused and troubled. Moreover, users will search according to their knowledge of terms, selecting familiar terms in preference to those that are unfamiliar. However, as we have seen, familiar terms are frequently multivocal (STIC's). Hence the natural priority in information retrieval for uninomialism unavoidably dictates multivocalism.

2.36 The Glossary Semantic. Scholars engaged in the production of knowledge, by contrast, assign priority to the need for unambiguous representation of their ideas. Accordingly, for each concept used in a subject field they need to have at least one univocal term, even though it be unfamiliar. It is this very unfamiliarity that leads critics of scholarly writing to criticize its technical jargon. However, once an intended concept has been clearly identified, writers may well prefer to use abbreviated terms, or synonymous terms that are familiar (even though multivocal).

If this can be done, in context, there is no need for the writer to repeat the use of unfamiliar and often lengthy univocal terms. Thus the scholar needs to be supplied not only with at least one univocal term for each of the core concepts of his/her subject field, but actually also requires a number of synonymous terms--making multinomialism a necessity.

2.37 The Choice of Semantic Modes. If the foregoing argument is valid, then one may well ask why glossographers typically select the inappropriate thesaurus semantic mode. To find the answer I believe one must find out who prepares glossaries. Typically glossographers are subject field specialists--they have no special knowledge or training in glossary design. Indeed, there is no recognized profession or specialty of "glossology," a field that lacks a parent discipline (like linguistics for lexicology, and information science for thesaurus design). Accordingly, (see #2.04) glossaries are typically designed according to the inappropriate paradigms provided by lexicographers and the compilers of thesauri. In order to test the validity of this position, the materials contained in the networking glossary have been analyzed and rearranged, as reported in Part II of this paper, and Annexes B to E.

Concluding Observations

2.40 To bring this theoretical discussion to an end, several additional observations will be offered. Although the difference between the needs of scholars engaged in the production of knowledge and those of information specialists and indexers involved in the retrieval of texts has been stressed, it is also important to understand their complementarity. There is no doubt that superior information services for scholars will help them be more productive. The reverse is also true: if scholars can be helped to write less ambiguously, then indexers will be able to improve the precision of their work also. Accordingly, the preparation of glossaries that are univocal/multinomial contributes to the improvement of retrieval services based on the contrasting principle of thesauri that are uninomial/multivocal.

2.41 Standardization. The unisemantic mode (see #2.32) enjoys some prospects of success in some fields of technology and natural science. It has been promoted by the International Standards Organization, through its technical committee on "Terminology." The field of Terminology has become identified with the effort to promote the use of a single preferred term for each concept, while also limiting the meaning of each term to one concept. Success in this effort requires that users modify their language habits to comply with a prescribed standard. It may well be that a variety of inducements can be used in these subject fields to promote such prescribed usages. In the social and information sciences, however,

efforts to prescribe preferred technical terms are unlikely to succeed.

2.42 "Neologisms." One reason for this resistance arises from the fact that familiar terms used for several concepts have to be abandoned in favor of neologisms, in many cases, if standardizing efforts are to succeed. As we have seen, however, social scientists are predisposed to use familiar words, as are information scientists, and to reject neologisms. This attitude rests in part on a misunderstanding of the meaning of the word, 'neologism.' According to Webster's unabridged dictionary, a "neologism" is any word or expression (whether newly coined or not) that is used for a new concept. This is precisely the meaning for which we have used, instead, the term, "semantic derivative." In addition, "neologism" sometimes means a newly coined word, and sometimes a meaningless expression used by psychotics! We shall do well to give up the use of the word, 'neologism," and to replace it with terms that more unequivocally designate the important distinctions that need to be made.

2.43 Words and Meanings. One of these distinctions has already been discussed (see #1.12) namely whether or not a given concept or sense is given in a general dictionary--i.e. the difference between semantic primitives and derivatives. We may add another distinction between words that are found in dictionaries and those that are not. Roughly speaking, the former are "old words" and the latter "new words," but these terms are unsatisfactory because, of course, some dictionary words are much older than others--hence one might argue that many "new words" are also found in dictionaries. In order to signify this distinction less ambiguously, we can use more technical terms. The word "neoterism" is not, despite its unfamiliarity, a "new word." It will be found in Webster's unabridged dictionary with a monosemantic meaning, namely that of a newly coined word. Instead, therefore, of using 'new word,' we could substitute 'neoterism.' Similarly, to designate a word found in general dictionaries, we could derive a new term based on the root, 'lex,' as found in 'lexical,' and 'lexeme." This word might be lexism* defined as any word found in a general dictionary as one of its entry words. Note that the word 'lexism' is itself a neoterism--it is not found in general dictionaries. If we combine the two criteria of semantic and lexical status, we can generate another matrix, as shown in Figure 3.

Figure 3: Terming Possibilities

		SEMANTIC STATUS	
		Primitive	Derivative
LEXICAL STATUS	Lexism	A	B
	Neoterism	C	D

2.44 Cell (A): Primitive Lexisms. The first cell in Figure 3 might be characterized as "primitive lexisms," i.e. words that appear in general dictionaries for defined senses which are also given in these dictionaries. In other words, all the terms found in general dictionaries are, by definition, primitive lexisms. If we think of the self-assigned task of lexicographers, we might see that they hunt for words of this type to include in their dictionaries--or we might say that, by including a word in a general dictionary, lexicographers give it the status of a primitive lexism. The other three cells in Figure 3 all raise questions of interest to glossology but not, I believe, to lexicology.

2.45 Cell (C): Primitive Neoterisms. The third cell in Figure 3 can be characterized as "primitive neoterisms," i.e. new words used for recognized concepts found in general dictionaries. It may be objected that if a concept that is defined in a dictionary already has a term, there is surely no need to coin a new word for it. Accordingly, one might think that there would never be any use for primitive neoterisms. However, if we consider the need of scholars for univocal terms, we will immediately see that if a term given in the dictionary has several senses, each of which is useful to the scholar, then it may indeed be useful to coin several new terms that can be used univocally for each of the relevant senses of a word. The example of 'text' has already been given, in #2.29. However, since 'text' in this sense is a semantic derivative, not a primitive, let us consider another example based on a primitive lexism, namely the word 'synonymy.' According to Webster's dictionary, this word has three senses: 1) a field of study; 2) a collection of synonyms; and 3) the quality of being synonymous. Let us suppose that all three of these senses are useful concepts, as indeed they are for anyone engaged in the preparation of glossaries. Although in many contexts it will be clear which sense is intended, in some contexts the word is used loosely and ambiguously. In order to make the distinctions more unequivocal, three new terms could be used, e.g. "synonymics" for (1); "synonyms" for (2); and "synonymity" for (3). Although 'synonyms' and 'synonymity' are obviously lexisms, 'synonymics' is a neoterism, and since it designates a concept found in a general dictionary, it is a primitive neoterism. From this example we can see that in glossary design it may be desirable to include not only new concepts and their terms, but also new terms for established concepts, i.e. primitive neoterisms.

2.46 Cell(B): Derivative Lexism. Turning now to the second cell in Figure 3, we shall immediately recognize that most of the terms used for new concepts in the special fields of knowledge of social and information science are derivative lexisms--i.e. they are words found in general dictionaries, but they carry new meanings. They are therefore always also polysemes, but they may be univocal or

multivocal, depending on how many concepts they designate within the subject field in which they are used. Clearly, in the design of glossaries for the social and information sciences, most of the terms presented will be derivative lexisms. Although it is correct, according to Webster's definition of "neologism" to use this word to include derivative lexisms, in popular practice most people think instead of derivative neoterisms--or even of primitive neoterisms--when they use the word.

2.47 Cell (D): Derivative Neoterisms. If social and information scientists persist in their present terming practices, we shall have the least need to use derivative neoterisms--i.e. newly coined words for new concepts. As we have seen, however, natural science and technology freely engages in the use of such terms, and this is related to the relative ease with which they accept standardized terms, as prescribed by official committees. We can now see that, apart from whatever practical incentives may be offered to users to use prescribed terms, it is also true that if a given user community is accustomed to accepting derivative neoterisms in their work, they are more likely to be receptive to terminological standards.

2.48 A Test. A simple test of the receptivity of readers of this paper to derivative neoterisms is offered in Figure 4. Let us assume that it will be useful to make frequent reference to the concepts identified in Figure 3, and that the compound terms, drawn from the defining characteristics of each of the four concepts discussed above (#2.44-2.47) prove awkward and inconvenient. We might decide that it would be useful to have, instead, a single short word for each of these concepts. It is not difficult to coin such words, using familiar roots. For the basic idea of early and late we can use 'eo-' and 'neo-'. For the idea of a word we can use 'nym,' as in 'synonym,' 'homonym,' and 'antonym;' and for the notion of a concept we can use 'log' as in 'logic' and 'logos.' This gives us the following synonymous terms:

(A) any word found in general dictionaries with an established meaning: primitive lexism, or eonym*

(B) any word found in general dictionaries with a new sense: derivative lexism, or eolog*

(C) any word not found in general dictionaries used for an established meaning: primitive neoterism, or neonym*

(D) any word not found in general dictionaries used to designate a new concept: derivative neoterism, or neolog*

Figure 4: Terming Neologs

		SEMANTIC STATUS: Primitive	SEMANTIC STATUS: Derivative
LEXICAL STATUS	Lexism	eonym	eolog
	Neoterism	neonym	neolog

2.49 Scholarly Preferences. The beauty of the approach to glossary design recommended here is that maximum freedom of choice is offered to users, while maintaining a high degree of dis-ambiguation. To use the concepts presented in ¶2.48 as an example, we have disaggregated the fuzzy set of ideas commonly associated with the word, 'neologism,' and established four significantly different, though closely related, concepts. Having done so, we have offered readers a choice of terms, e.g.

new word for a new concept	old word for an old concept
derivative neoterism	primitive lexism
neolog	eonym

Anyone interested in making the conceptual distinctions examined in ¶2.48 may, of course, choose to use all of the available synonymous terms, noting that the shortest and most convenient are also the most unfamiliar, being completely new coinages, whereas the longest and most cumbersome are also the easiest to remember and to learn. The terming "neologs" that have just been introduced will not again be used in this essay--the concepts they designate are not central to the design of glossaries, as are the concepts of uninomial/multinomial, and of univocal/multivocal, and the concepts of tracking, term replacement, and elasticity. The purpose of this concluding exercise, then, has not been to add another set of concepts so much as to provide an experience that would help readers understand the reasons why terminological standardization is not a readily acceptable approach for specialists in the social and information sciences, and why the presentation of sets of synonymous terms offers, instead, a more acceptable paradigm for the design of glo-saries--at least in the social sciences.

THE DESIGN OF GLOSSARIES: An Example from "Networking"

by Fred W. Riggs, University of Hawaii

PART II: TOWARD SOLUTIONS

3.00 The problems of tracking and of polysemy/synonymy as they relate to the design of glossaries will be illustrated with a view to finding solutions, in the material that follows. The theoretical discussion in Part I of this essay will be utilized as a framework for the analysis. Materials contained in the Glossary for Library Networking, published by the Library of Congress in 1978, provide the data for this presentation. They will be handled in a series of annexes follows:

B. a tracking exercise, using 'sink' and 'source' as starters;
C. a chained arrangement of the records generated by the tracking exercise;
D. a transitional arrangement of these records in classified order;
E. a revised classified arrangement; and
F. an alphabetical index of the terms used in Annexes B to E.

An additional annex, G, provides an index for new concepts relating to the design of glossaries that were introduced in Part I; and Annex A, which was attached to Part I, gives some sample pages taken from the networking glossary.

3.01 'Sink' and 'Source'. The tracking exercise identifies all of the defined terms that must be understood in order to comprehend two selected definitions: those of 'sink' and 'source.' Although this material was selected primarily to illustrate the tracking problem, it also demonstrates how dependence on the thesaurus semantic mode leads to polysemantic ambiguity as the price for uninomialism. (see #2.2- in Part I)

3.02 Alternative Designs. The remaining annexes (following B) illustrate possible alternatives to the alphabetical thesaurus-based arrangement of records that was used in the networking glossary. Annex C provides a chained sequence which maximally simplifies the solution to the tracking problem, and produces a re-arrangement of the records identified in Annex B. Such an arrangement of glossary records might well be used in a primer designed to teach networking to beginners. As an intermediate stage in the production of glossaries, it could also be used to unravel definitional circularities that easily creep into the defining texts of a glossary. The classified arrangement of records illustrated in Annexes D and E puts the material in an order that facilitates comprehension of the various meanings that have been distinguished from each other--and also helps us compare definitions that are virtually equivalent, making distinctions without significance. Although these arrangements also help to solve the tracking problem, they give higher priority to the issues of polysemy/synonymy.

3.03 Revisions. Annex D uses only terms and definitions that have been taken, without modification, from the networking glossary. This procedure highlights the lack of univocal terms for many of the concepts given in the glossary. One way to overcome this limitation is illustrated in Annex E. These examples show that, in order to supply univocal terms for all of the concepts used in networking, it is necessary to have more than one term for many of them, simply because the more familiar terms preferred by users for some of the concepts are multivocal. Illustrative revisions of the defining texts are also given in Annex E. They result, in part, from the replacement of multivocal with univocal terms, and in part from the logical interdependence of definitions in a classified array.

The Tracking Exercise

3.10 By tracking the defined terms used in the definitions of 'sink' and 'source' we can illustrate some of the problems that occur when interdependent definitions in an alphabetized glossary are tracked. Although these two words were not quite selected at random, other terms could just as well have been chosen and would have given similar results.

3.11 Tracked Terms. Annex B provides information on the outcome of the tracking exercise. The defined terms, taken from the networking glossary, draw in records not included in Annex A. The material is arranged by "steps," each of which identifies the entailed terms whose definitions must be supplied in order for the concepts in the previous steps to be understood.

3.12 Marked Terms and Sleepers. Although most of the tracked terms are marked, some are sleepers. (see #1.25) The marked terms are underlined in the original text, and so may easily be found. As for the unmarked terms, most of them are no doubt semantic primitives, although one cannot be sure that words like 'computer,' 'processes,' and 'connected' are not semantic derivatives. However, the words, 'data,' 'network,' and 'node,' are defined in the glossary even though they are unmarked when used as entailed terms. To remind readers that these words were sleepers in the original glossary, they are underscored with a broken line, in Annex B, and they have been listed in parentheses in the list of entailed terms on the right hand margin of the annex. However, the distinction between sleepers and marked terms is not carried into the subsequent annexes, where the sleepers are underscored in the same way as the defined terms that are marked in the original publication.

3.13 Steps 2 to 7. The identification of the seven steps required to reduce the definitions of 'sink' and 'source' to ordinary language terms (semantic primitives) is supplied by the tracking terms given in the right hand column of Annex B. Each of these terms is given only once, following its first use in a defining text, and the number after each of these terms indexes its location in the next step. At the 7th step we come to a defining text all of whose entailed terms are either primitives or have been defined in previous steps. To reduce the definitions in this set to ordinarly language terms would, of course, be a monumental task, since each of the entailed terms that is defined in the glossary would have to be replaced by its definition, generating impossibly lengthy stretched out definitions for, let us say, 'sink' and 'source.' In general, the more steps required to track a definition, the more inelastic that definition is--and the more important it is,also, to make the task of tracking simpler than it can ever be in an alphabetized glossary. (see #1.37)

3.14 User Experience. No doubt experienced users, who already are largely familiar with the vocabulary of library networking, will find that an alphabetized glossary is convenient to use, inasmuch as they will rarely have to track a definition because they will be familiar with most of the technical terms that it contains. By contrast, outsiders and beginners who seek to penetrate the mysteries involved in a new subject field will quickly become discouraged by the effort required to track definitions to their roots in ordinary language terms.

3.15 The Perils of Sleepers. Additional tracking difficulties for newcomers to a field are caused by the presence of sleepers in concept definitions. Unsuspectingly they may assume, because these defined terms are unmarked, that they are used as primitives, i.e. in accordance with their general dictionary definitions. Although only three sleepers--'data,' 'network,' and 'node'--are found in the sample presented in Annex B, they are employed in 19 (or 70%) of the 28 concept definitions used in this annex. Several definitions actually contain two or more sleepers!

3.16 Ambiguity from Multivocalism. In the tracking procedure, the risks of ambiguity become apparent. The first definition of 'sink,' for example, uses 'terminal' and 'channel.' When we look up the definitions of these terms, we find that they are multivocal, since each can be used to designate two different concepts within the field of library networking. Because we cannot easily determine which of these senses is

intended in the definition of 'sink'--and the same problem occurs in the definition of 'source'--ambiguity based directly on multivocalism is uncovered. Since, as we shall see, the multivocalism of this glossary arises from the priority assigned by its authors to uninomialism, we could also say that ambiguity in the glossary is due, ultimately, to the insistence of its authors on using only one preferred term for each concept.

3.17 The Extent of Ambiguity. The networking glossary as a whole contains 171 concept definitions, but only 152 defined terms, which means that 19 of the terms are multivocal, or one-eighth of the total number. However, they are assuredly among the most frequently used terms, if we may judge by our sample. Among the records given in Annex B, we find that 12 of the 28 concept definitions contain multivocal terms, or almost half of the total number. Accordingly, anyone not already familiar with the field of networking could easily be confused by the glossary when attempting to construe which of the possible meanings of these terms was intended. This source of ambiguity can be readily overcome by the use of more than one synonymous term for the concepts whose terms are multivocal. In other words, by giving up the high priority for uninomialism it becomes possible to achieve univocalism.

3.18 The Preference for Uninomialism. We have already explained why uninomialism, in the thesaurus mode, is important for retrieval purposes, and why, by contrast, univocalism is more useful to scholars engaged in the production of knowledge. (see #2.33 and 2.34) The consequences of the preference for uninomialism when used (inappropriately) in the design of glossaries can be seen in the data presented in Annex B. For all but one of the concepts identified in this sample, only one term is supplied. The exception is the concept defined in (3d) for which we have two terms, 'network node,' and 'node.' (3e) The latter is given in a separate cross-reference only, and is not repeated as a synonymous term in the main entry for the concept. Users, therefore, will not automatically learn, from the entry for 'network node' that 'node' can be used as a synonymous term. This is indeed paradoxical since we find that of the five definitions contained in Annex B that make use of this concept, four actually use 'node,' (2c, 3a, 3b, 3c) whereas only one (4a) uses 'network node.'

3.19 The Suppression of Synonymous Terms. In the glossary as a whole, there are 171 defined concepts, plus 30 "see" references, indicating that less than one-fifth of the concepts included in the glossary are supplied with more than one term--and then only in separate entries, without any "also use" references to the synonymous terms in the main entries for each concept. Although I have no expertise in the field of networking, I would be surprised if many more synonymous terms for these concepts were not in use among practitioners. The result is--if my assumption is valid--that users of the glossary are not informed about the synonymous terms that have been used, nor about their context of usage, i.e. who or what group employs each term for the defined concepts. Even the few synonymous terms that are given in the cross-references are not called to the user's attention. The glossary, then, on the basis of the thesaurus semantic mode, is designed to constrain users to employ only one term for each concept, even though that term may have several meanings within the field of networking. Useful as this may be for retrieval purposes, it is unhelpful to those engaged in the production of knowledge about library networking--or of any other subject field.

The Inverted Entry (Record)

3.20 In order to overcome some of the difficulties encountered in the networking glossary, it is useful to put glossary records in the format of an inverted entry, i.e. to start with the defining text and to give the term family (glossonym) for each concept after this text. This is not to say that the usual order of a dictionary entry (term-to-definitions) cannot be followed, but only that the pressures of conventional models would almost certainly be so great as to make it appear that the first member of each term family should be a preferred term, one that could be used to alphabetize the records. The inverted order enables us to overcome this bias.

3.21 Glossary Records. Those who think of glossary records on the analogy of dictionary entries are likely to rebel, almost instinctively, against this procedure. It is precisely for this reason that the word 'record' has been used in preference to 'entry.' In a non-alphabetical glossary, after all, records are not entry points--one has to rely on the alphabetical index to supply the "entries" which guide users to concept definitions, via notations or serial numbers. A record, therefore, is precisely not an entry point, and it is essential to stress this distinction by considering, here, the format of catalog records.

3.22 Catalogs. A record format that is familiar to all may be found in any merchandise catalog. The Sears Roebuck Catalog, to mention a familiar example, consists of a classified array of product descriptions, accompanied by catalog numbers and labels for each product. To facilitate the locating of desired products an alphabetical index is supplied. However, the classified arrangement of products also permits the collocation of similar products so that anyone looking for different kinds of clothing, furniture, camping equipment, or kitchen ware would be able to find their descriptions. This kind of "browsability" is an important feature of any classified arrangement of items--including concept definitions, or books on a library's open shelf collection.

3.23 Atlases. Similarly, in an atlas the main "text" consists of maps on which geographical place names are plotted in their spatial settings. Anyone travelling about or seeking to visualize a place on the map needs to see its name in context with other locations. An alphabetical index, of course, is a necessary adjunct of any atlas, permitting one who knows the name of a town, river, or mountain to find out where it is located. Primary emphasis in an atlas, as in a catalog, is placed on the contextual location of items, not on the alphabetical arrangement of their names.

3.24 Directories. A city directory (by contrast with telephone directories) is also arranged according to a classification of locations, rather than by the alphabetical ordering of names. Although there are a wide variety of reference works arranged by classification schemes, such as those found in catalogs, atlases, and city directories, we have no generic name for them that can be used to designate the inverse paradigm of an (alphabetized) dictionary. Yet this is precisely the model we require for our work on the design of glossaries.

3.25 The Glossary Record Format. If we think of the records found in glossaries on the analogy of those found in catalogs, atlases, and directories--rather than of dictionary entries--we shall see that it is reasonable to ascribe the following elements to each such record, in this order (or format):

a) Notation: first a number, like a catalog or class number, (or serial number?) which not only locates each record in a meaningful sequence, but permits ready access via the index.

b) Defining text: second, a text which sets forth the defining characteristics of the concept in question. It is important not to think of the defining text as though it were the "definiens" of a definition in which the thing to be defined, the "definiendum," is a word. It is important, however, to mark clearly the semantic derivatives that are used in each defining text so that users will easily learn which words have meanings not given in their general dictionary definitions. A further distinction should be made between the marked terms that designate core concepts of the glossary's subject field, which are defined elsewhere in the glossary, and other semantic derivatives whose definitions have to be searched for at some other place.

c) Glossonym: third, a set of synonymous terms for each concept identified by the defining text should follow next. The members of each such glossony should be separated into two sub-sets, those that are univocal and others that are multivocal.

i. "STOC" It is convenient to use the label, "STOC" before the univocal terms to show that, in the context of the glossary's subject field, they designate only one concept--however many other meanings they may have in other fields.

ii. "STIC" Multivocal terms may be preceded by the label, "STIC" to show that each of them can be used for more than one concept within the subject field of the

glossary. Of course it is arbitrary to use 'STIC' and 'STOC'--any other synonymous labels could be used provided these meanings are clearly indicated.

d) Sources and Contexts. Additional information should be provided, notably the bibliographic source for each defining text, and context labels for the terms included in each glossonym. Since this information is not supplied in the networking glossary, it cannot be given in the attached annexes. Context labels may take the form of reference numbers for bibliographic sources, thereby showing the user which authorities prefer each of the synonymous terms. Of course, the authors of a glossary may also introduce their own preferences, with suitable credit markers. The important point is simply to let users know who is responsible for each of the concepts and their terms. Glossaries may also facilitate multilingual translation, notably by including terms in several languages, again with markers to show the language in which each term is used. (It would probably be a mistake, however, to offer defining texts in more than one language in a single record. A better practice would involve the preparation of parallel records in different languages--i.e. one in English, another in French, a third in Russian, etc.)

3.26 Examples: Chained and Classified. In order to illustrate the principles set forth in Part I, using the record structure proposed in #3.25, the material presented in Annexes C, D, and E will now be discussed. The data elements used for these presentations are those contained in Annex B. Since all of the concepts identified by tracking the definitions of 'sink' and 'source' in the networking glossary are, in principle, interdependent, it is convenient to treat them as a linked set, and it turns out that they provide enough data to provide the intended demonstration. Two organizing principles will be used: 1) the chaining principle, as illustrated in Annex C; and 2) the classification principle, as shown in Annexes D and E.

Chaining Records

3.30 The method of record chaining involves concept derivation, the inverse of the method of reduction involved in term tracking. As we saw in Annex B, tracking requires one to hunt for the definitions of entailed terms until, finally, one arrives at definitions that entail only the use of primitive terms. (see #1.14-17, and 1.31-34) The inverse procedure involves starting with definitions that use only primitives, and then step by step using the terms that have been so defined as elements in the definition of concepts which follow. These procedures have been explained in Part I: the procedure of reducing definitions to semantic primitives parallels the procedure of tracking (see #1.32-34); and the procedure of concept derivation involves building from semantic primitives to derivative concepts (#1.14-17). A set of chained records illustrates the procedure of concept derivation; while a set of tracked records permits the reduction (stretching) of definitions. In this sense, the content of Annex C is an inverse of the content of Annex B, even though the sequence of records is not necessarily reversed.

3.31 The Starting Point. In order to initiate a sequence of chained records, then, one must find starting concepts which are primary derivatives, i.e. their defining texts contain only semantic primitives, terms that can be understood by reference to one of their senses as given in a general dictionary. Although we cannot assume that any set of interdependent concepts, such as those given in Annex B, will contain primary derivatives, inspection reveals two, namely 'data' and 'link'--provided we agree to drop the term, 'node,' from the definition of 'link.' (see #3.34)

3.32 Notations. In order to facilitate the indexing of records in the different Annexes, we will not use the same notation scheme in each: thus records in Annex C are assigned two-digit numbers; whereas the records contained in Annex B have a number-letter combination; and the records in Annexes D and E use three-digit numbers, each pointing to the same concept in the two annexes.

3.33 Pursuing the Chain. Having established the meaning of 'link' and 'data' as primary derivatives, we can proceed to define secondary derivatives in which these terms

can be used, i.e. 'channel,' 'control function,' and 'telecommunications.' Knowledge of these three secondary derivatives now permits us to advance to a third level, at which the meanings of 'station' and 'control character' can be established. Following this procedure, we discover that it is possible to define all of the terms given in Annex B in a chain of derivatives that proceeds for eight steps.

3.34 Uses of Chained Records. We may now ask if there is any use for chained records. Quite apparently there is no apparent logic in the ordering of concepts, apart from the levels of semantic derivation. Yet a glossary containing chained records might be useful as a learning tool. If attached to a primer, the chained series of records could be linked with a learning sequence in which the beginner proceeds from the known to the unknown, from the familiar to the unfamiliar. Another possible use is suggested by the use of 'node' in the original definition of 'link.' (see 3.31) It helps us identify circularity in definitions. Let us consider this point more closely.

3.35 Circularity in Definitions. If we observe the logic involved in the following reduction (tracking exercise) the point may become more evident:

(1) link: a communications path between two nodes or points

(2) node: a...terminal, or communications computer in a computer network

(3) network: an interconnected or interrelated group of nodes

(4) terminal: a device for entering data into ...a computer network

(5) communications computer: a special purpose computer used to control or format data transmitted between network nodes

Consider the following points: 1) to understand the meaning of 'node' we must know the meaning of 'network,' but in order to understand 'network' we need to know the meaning of 'node'; 2) an understanding of 'link' requires prior knowledge of 'node,' but to gain this knowledge we must know what a'network'is and this, as we have seen, presupposes a knowledge of 'node'; 3) and this applies to other concepts in which 'nodes' occur as entailed terms, such as 'communications computer,' since to understand (5) we require a prior knowledge of (2), but (2) presupposes a knowledge of (5). In order to by-pass this circularity, we dropped 'node' from the definition of 'link' given in Annex C (12), arguing that 'points' could be treated as a synonymous term. However, inadvertently the definition of 'node' was given, at (64) in Annex C, before the definition of 'communications computer'--but the error could not be rectified because even if the definition of 'communications computer' were given before the definition of 'node,' it would have presupposed a knowledge of the meaning of 'node.' This unavoidable error in the chaining of Annex C stems from an intrinsic circularity in the original glossary--a circularity that would probably not become evident save for this kind of chaining exercise.

3.36 An Intervening Stage. One conclusion to be reached from the chaining shown in Annex C, therefore, is that even if one might now want to use or publish a glossary containing chained records, it might still be useful to prepare a set of records in this form so as to uncover hidden circularities that impair the internal logic of the concept definitions.

Classifying Records

3.40 In addition to chaining records, another non-alphabetical procedure involves the use of a classification scheme. Concepts can be ordered so as to reflect their relations to each other. One of the objections sometimes raised to the classification of concepts in a glossary arises from the fact that a wide variety of classification schemes could be used. How, one might ask, can we know which is the best system for classifying concepts? A scheme that is acceptable to some users may be unacceptable to others. It should be stressed immediately that although the theory of glossary design leads to emphasis on the need for classification, it by no means dictates any particular scheme or system. Rather, priority has to be given to the preferences of users: they should feel free to select any classification scheme that they are happy with.

Moreover, if different members of a user group prefer more than one scheme, there is no reason why more than one classification should not be used, any more than it is necessary to choose only one term per concept. By means of computerization, it is easy enough to generate a printout in which records can be ordered according to a variety of schemes, all of which can be combined in a single publication, just as the separate Annexes to this paper display different ways of arranging the same set of concept records. The alphabetical index of a glossary can point, with equal facility, to the location of every concept record, in each cloassification scheme--just as our index in Annex F does.

3.41 Frequency of Use. Many users will, no doubt, want to start with one of the existing universal classification schemes--such as Dewey or L.C. in the U.S., the U.D.C. in Europe, Bliss, Colon, Mison Rubricator, or some completely new system, such as the one designed by Dr. I. Dahlberg, editor of International Classification--and attempt to fit the new concepts of their field into their preferred scheme. Another approach would start more simply by identifying the subject field of the glossary according to an established system, such as the new FID/BSO, which has been sponsored by UNESCO for precisely this purpose. If such an approach were to prove acceptable, it would enable users of a glossary to tell at a glance where, in relation to other subject fields, the field in question may be located. Subject to a field code or notation, e.g. within the BSO system, all class numbers might be designed by the authors of a glossary in conformity with their own preferences and understanding of the structure of their field of interest. A good starting point for such an analysis might be supplied by a content analysis of definitions. If we assume that the use of a term in a defining text usually shows that the defined concept is relatively specific--i.e. that the entailed term designates a broader concept than the defined term--then we might want to think about the most frequently used terms as possible starting points. In the records contained in Annex B we find that the following terms appear most often:

term		frequency
'data'	found in:	16 defining texts
'network'		10
'node'/'network node'		4
'message'		3
'terminal'		3
'header'		3

The use of these words as entry terms was not counted, nor were they counted when used in "see also" references, or in explanatory comments that follow a definition. Study of these terms provides some instructive clues. For example, the word 'data' suggests that one of the main categories could be the content of messages, and the frequency of 'network' indicates the importance of system structure. Would it be appropriate, therefore, to use a bifurcated classification scheme containing two main categories: 1) the contents of information transmitted in a network; and 2) the architecture or machinery used to carry these messages?

3.42 Multivocal Terms. If it were not for the presence of multivocal terms in the networking glossary, the answer to the preceding question might well be "yes." However, we have to take into account the fact that some of the most frequently used terms designate more than one concept. Among the frequently used terms listed above, three (50%) are multivocal. If we compare the two meanings assigned to each of these terms, we find that one refers to a structural or architectural feature, but the other designates some kind of process or function. The point is more easily grasped if we construct a simple array, as follows:

	term	structure	process
(6)	'network'	a group of nodes	pattern of information exchange
(7)	'terminal'	a device for entering data	a point where data can enter or leave
(8)	'message'	a single transmission	a unit of transmitted information

A similar distinction separates the two meanings of three other terms used in our sample: 'channel,' 'source,' and 'sink.' When we assess the significance of these multivocal terms, we see that, perhaps, a more generic distinction has been made in the set of concepts defined in Annex B: namely, the structure/process dichotomy. If we add this distinction to the dichotomy drawn from the term-frequency analysis given in #3.41, we arrive at a possible three-fold categorization, as follows:

1) the architecture (or structure) used in networking;

2) the contents transmitted by this structure; and

3) the functions (operations, activities) of those who use the structure.

Specialists in library networking may well protest that a different classification would be more useful--since I am not a specialist in this field, I could only defer to their preferences, but add that several different classification schemes can also be used, together. It might also be objected that the sub-set of the networking glossary's terms derived by tracking the definitions of 'sink' and 'source' are by no means representative of the whole set. I can only agree, and ask specialists in the field of library networking to revise and complete the exercise presented here, for illustrative purposes only. Actually, I think it is rather amazing that the small sample collected by a tracking exercise lends itself so well to a simple classification.

Annex D

3.50 The terms found in Annex B have been sorted into three main categories, as indicated in #3.49, and then again sub-sorted into sub-categories--but I shall not try to explain the considerations that led to the results presented in Annex D. Let me rather explain the format of the annex, which is based on the prescriptions for a record format, as given in #3.25. This format gives us four columns, as follows:

1) the notation (consisting of three digit numbers);

2) the defining text;

3) the univocal terms (STOC's), and

4) the multivocal terms (STIC's).

Since bibliographic and contextual information is not included in the networking glossary, it could not be added here. The third and fourth column, combined, form the "glossonym." Sense numbers have been excluded in order to provide a more realistic experience of the characteristic ambiguity entailed by using STIC terms. If more than one term is given in the glossary for a single concept, both are included--only one instance occurs, however, in the data from Annex B--namely, 'node' and 'network node'--but more would occur if the whole glossary could be analyzed in this way.

3.51 The Tracking Problem. If we evaluate the material presented in Annex D from the point of view of tracking, we shall see that it does not simplify the problem's solution so much as the chained glossary does, but it is indeed easier to track definitions in the classified records than it was in the alphabetized arrangement. If we suppose that a user starts from the beginning to read through a classified glossary, we will find that ten of the defined terms found in the first category (the 100's) have to be tracked. Half of these, however, can be located near by, i.e. in the same category, but the rest must be searched for in the subsequent parts of the scheme. Moreover, the tracking task becomes simpler as one proceeds. If we assume that a user will master the terms defined in the first category before proceeding to the second, then he or she will have to track only four of the 12 entailed terms used in the second category, and none of the marked terms used in the third category. To make a direct comparison with the problems involved in tracking the meaning of 'sink' and 'source' from the classified glossary, as given in Annex D, clearly the meanings of these two terms can be established far more easily than they could if one had to use the procedure illustrated in Annex B.

3.52 The Polysemy/Synonymy Dilemma. When we turn to the question of concepts per term and terms per concept we can see, from Annex D, that the networking glossary does not assign univocal terms to 13 (almost 50%) of the 28 defined concepts given in our sample. The reason for this situation now becomes quite apparent: if only one term (uninomialism) is to be assigned to each concept, then the familiar terms in use often have to be used for more that one concept, as may readily be seen by reference to the terms listed in the column under "STIC." Moreover, since these seven words are used 23 times as entailed terms, without any indication of the intended sense, they contribute substantially and unnecessarily to the ambiguity of many definitions. Quite apparently, then, the use of a thesaurus semantic mode, which permits multivocalism in order to achieve uninomialism, has led to definitional ambiguities. This difficulty can easily be rectified by supplying a univocal term for each concept, while retaining the multivocal synonyms. A procedure for doing this is illustrated in Annex E. (The reader must remember that the univocal terms supplied in Annex E are not intended as recommendations. Assuredly, specialists in library networking have the right and the duty to select their own terms. Rather, the illustration is intended only to show how simple it is to design univocal terms by the addition of qualifying words--and other techniques are also available.)

Annex E

3.60 The concept records contained in Annex E are, essentially, the same as those given in Annex D, with some small changes to be discussed below. The most important change involves the addition of arbitrary univocal terms for the concepts that lack them--they are given in parentheses, under the "STOC" column. In addition, the definitions in which multivocal terms are used have been revised by the replacement of the new univocal terms. In addition, the classification of concepts highlights both possible gaps and redundancies in the materials, and has led to the addition and subtraction of records. Some definitions have also been condensed by the substitution of defined terms for explanatory phrases, a procedure that is more acceptable when the definitions of the defined terms are easily tracked.

3.61 The Use of Sense Numbers. The simplest way to create univocal terms is to use sense numbers. By this means, we can distinguish between 'network(1)' and 'network(2),' for example, However, this device gives us no mnemonic assistance--we may find it difficult to remember which meaning of 'network' is no.1, and which is no.2. To visualize this solution and the difficulties that it also creates, consider the following illustrations:

(9) network(1): two or more organizations engaged in a common pattern of information exchange...

(10) network(2): an interconnected...group of nodes

(11) communications network: the physical means for a group of nodes to intercommunicate data

(12) source(2): the point of entry of data into a network(1)

(13) terminal(1): a device for entering data into...a computer network(2)

Actually, I can only guess that the intended sense of 'network' in the definition of 'source(2)' is actually 'network(1),' or that the sense of the same word as it is used in the definition of 'terminal(1)' is 'network(2).' Since no sense numbers are used in the networking glossary, I may have garbled its authors' intentions. However, I suspect that even if such numbers had been used, it would have been difficult for users to remember them--as a test, try to remember (9) and (10) when reading (12) and (13). Note also that, although (11) is a kind of network, we are enjoined not to use this word for the concept, even if its meaning in context is unambiguous.

3.62 The Use of Menmonics. An alternative procedure involves the creation of mnemonic terms, in which descriptive adjectives, the names of authors, or other qualifying words are used in order to distinguish between the various senses of a

word--sometimes quite different expressions can also be used. For example, in Annex E (322) I have suggested the use of 'communication' as a univocal term that might serve as a synonym for 'message(2).' All of the other examples involve the addition of qualifying words, however. To illustrate, consider the following examples, all of which involve paraphrases of (9) to (13).

		STOC	STIC
(14)	organizations using a communications-network	functioning network, network(1)	network
(15)	an interconnected group of nodes	network structure, network(2)	network
(16)	a network structure used to communicate data	communications-network	(network?)
(17)	a point of entry of data into a functioning network	functioning-source, source(2)	source
(18)	a device for entering data into a computerized network structure	terminal structure, terminal(1)	terminal

The following term replacements have been used:

'functioning network,' for 'network(1)'
'network structure,' for 'network(2)'
'functioning source,' for 'source(2)'
'terminal structure,' for 'terminal(1)'

In addition, one might consider using 'network' in a third sense, to designate the concept of a 'communications network.' In practice, of course, writers would want to make sure that it was complete clear to readers which of the several possible meanings of 'network,' (i.e. 14, 15, or 16) was intended, and if there was any possibility of ambiguity, the appropriate univocal (STOC) terms given above could be used instead of the simpler, but multivocal, (STIC) term.

3.63 A Mnemonic Test. As a test of the utility of the mnemonic terms given above, one might compare the definitions given in #3.62 with those provided in #3.61. Start by comparing (17) with (12), and then (18) with (13). Is it easier to remember the meaning of 'network(1)' in the former, and 'network(2)' in the latter (as given in (12) and (13), or the terms 'functioning network' and 'network structure.' As for (16), the expression 'physical means for a group of nodes' has been replaced, by 'network structure.' This replacement not only enables a user to grasp the difference between concepts (15) and (16), but also facilitates the use of 'network' in a third sense, since it becomes much easier to differentiate the meanings (three?) of the term in library networking in case there should happen to be any possible ambiguity. The point of the exercise is not to require more cumbersome terms in place of simpler ones, but only to enable writers (users) to show, in a very easy way, which of the several possible meanings of a much-used and short expression is intended. Either sense numbers or mnemonic terms could be used for this purpose, but my hunch is that the latter will be more useful--of course, both can also be used.

3.64 The Example of 'Text'. To take another example, consider the use of the word 'text.' The three senses of text given in Annex D, at 233, 221, and 235 are surely hard to remember, if designated univocally by their sense numbers, e.g. as 'text(1),' 'text(2),' and 'text(3).' We might, instead, use such mnemonic terms as, 'whole text,' 'message text,' and 'running text.' To clarify the distinctions that might be made, let me paraphrase the definitions supplied in the glossary as follows:

(19) all the elements included in a message, including control characters: whole text

(20) all the (meaningful) data elements included in a whole text: message text

(21) the data elements that follow a header in a message text: running text

When we attempt to match these paraphrases with the actual definitions given in the text (see Annex D, 233, 221, and 235) we encounter a difficulty, because we find that the definition of 'text(2) (221) means a kind of "format." One would think that there is a useful distinction between the contents of any text, and its format. If so, there may be need for both a set of format concepts, and for a set of corresponding content concepts. To clarify this distinction, we might construct a simple table, as follows:

		content	format
(22)	all elements in a transmission	whole text	whole text format
(23)	data elements in a message	message text	message text format
(24)	data elements following the header in a message text	running text	running text format

If we now compare the definitions shown in Annex D with these categories, we can see that 'text(1)' identifies the sense of 'message text,' and 'text(3)' the sense of 'running text.' However, 'text(2)' appears to define the sense of (22) or (23). The definition of 'text(2)' contains a "see also" reference to 'message text format.' This suggests a related, but not a synonymous, term. The examples given in both definitions include, 'query text,' 'reply text,' and 'error text,' but the example under 'message text format' includes 'header length,' and 'data length,' which are not meaningful to the user, and hence may be 'non-data.' It appears, therefore, that the definitions of 'message text format' and of 'text(2)' could both be construed as designating either 'whole text format' or 'message text format.' Moreover, there is no record in the glossary that unequivocally defined the sense of (22). To cope with this problem, a glossographer would have to raise questions for specialists in the field to answer. One approach would be to design a rough draft, indicating a possible solution, permitting users to decide how they wanted to have it resolved. In this instance, again for illustrative purposes, I have inserted into Annex E an extra record (213) to accomodate the concept of 'whole text,' and have revised the definitions of (221) and (222) so as to make them distinguish clearly between the 'whole text' and the 'message text' levels. In both instances, of course, they designate a format, and the juxtaposition reveals that there is no concept of a "running text format" in the glossary--but it may also be unnecessary to add such a concept if it cannot be partitioned into significant components.

3.65 Re-Definitions. The examples just given suggest the utility, in glossary analysis, of proposing re-definitions of concepts. Several examples will be found in Annex E. One type of re-definition has already been discussed, namely the replacement of a multivocal term by one that is univocal, e.g. the substitution of 'network structure' for 'network' in #133—(13) and (18) above. Sometimes condensation may be used, as illustrated above in the shift from (11) to (16)--see Annex E, no.112. Sometimes more far-reaching changes appear to be justified by the logical structure of interrelated concepts. This can take the form of additions and subtractions, as indicated in #3.64, in the handling of 'text(2)' and 'message text format.' Similarly, the logic of "representations to which meaning can be assigned" (211) implies an opposite concept, namely "representations to which meanings are not assigned,' leading to the addition of a new record for 'non-data,' (212 in Annex E), a term whose utility becomes apparent in relation to the meaning of 'control character'--see #240 and 241. Working in the reverse direction, a comparison of the defining texts of #231, 232, and 233 suggests that no important distinction has been made--or at least the distinction is not apparent to me. Perhaps if the definitions were to be revised, the significant differences between 'message(1),' 'text(1),' and 'message text' could be clarified. To me, they appear to be synonymous terms, and the defining texts make distinctions which I do not understand. Accordingly, in Annex E, I have replaced the three records with one, i.e. #231. This does not mean, of course, that there are not important distinctions to be made in this matter--perhaps they should be added. Another example occurs in the definitions of 'data' (#121) and 'data link' (#122), which puzzle me. Since I

cannot understand the significance of the difference, I have collapsed the two records into one, #121 in Annex E. Perhaps specialists would want to make a more evident distinction, and they could easily add another concept, if this were an actual glossary development exercise. Ideally, as noted elsewhere, the development of a glossary should be a continuing process of interaction between specialist users, and a few glossographers--in any lively subject field, continuous developments and innovation require continuing revisions and additions in the set of new concepts and their terms with which a user community is working.

3.66 A Glossary Semantic Mode. If our exercise has been successful, we may see in the way the materials in Annex E have been presented a solution for the key problems of glossary design. The inherent difficulties encountered when tracking the interdependent definitions of an alphabetically arranged glossary have been substantially reduced, even if not to a minimum. The chained record format, shown in Annex C, would indeed reduce the tracking problem even further. However, the logic of interrelated concepts becomes more evident in the classified structure used in Annex E, and this appears to be a more important consideration. Moreover, the use of this mode facilitates the tightening of definitions and the provision of a term family which, in every instance, supplies a univocal (STOC) term, without inhibiting the employment of simpler multivocal (STIC) terms whenever their use, in context, is unambiguous. Users following the guidance provided by such a glossary would be enabled to designate clearly the technical concepts they have in mind, while enjoying the stylistic options and simplicity of expression that the availability of synonyms permits. The classified glossary thus facilitates a solution to the problems of polysemy/synonymy in a manner appropriate to the needs of scholars engaged in the production of knowledge--including scholars writing in the fields of information science. Without taking exception to the essential virtue of the thesaurus semantic mode (uninomial and multivocal) as an approach suitable for the indexing and retrieving of completed documents, it is important to understand the importance of the alternative glossary semantic mode (univocal and multinomial) as the one that best meets the needs of scholars engaged in the writing or production of documents.

Indexing

3.70 To say this, however, is not to say that an alphabetical list of terms is not necessary. However, if the main body of a glossary consists of a classified set of records, then an alphabetical index is clearly required as a supplement. This is actually an advantage because all of the synonymous terms used for a given concept can be listed in an index without the need to make invidious distinctions--it is not necessary to select one of them for a preferred status, as it is when a single term is chosen to serve as the main entry of an alphabetized array. That practice in alphabetized glossaries is dictated by considerations of economy, since it is cheaper to avoid repeating definitions at several locations.

3.71 Synonymous Terms. Although the synonymous terms for each concept included in a classified glossary can be included in its glossonym, it is helpful to add them to the index, immediately after each index term, as illustrated in Annex F. This is not a necessary feature, admittedly, but if it can be done, it helps the user who may know one of the terms for a concept, and may not need to look up its definition.

3.72 Notations. Several classification schemes may be used in a single glossary, as noted above, #3.40. If each scheme is marked separately, they can be identified with ease in the index, so that users will quickly find the concepts they are looking for in the preferred classification scheme. An analogy may be mentioned: published catalog cards typically give at least two class numbers, one for the Library of Congress classification scheme, and a second for the Dewey numbers. It is, of course, unnecessary to mark one "LC" and the other "Dewey," since the form of the notations marks the difference, the former being alpha-numerical, and the latter numerical. A similar distinction is made here by using alpha-numerical notations in Annex B, two-digit numbers in C, and three-digit numbers in D and E. An index, in short, can point by its notation conventions to as many different classification schemes as are required.

3.73 Textual Indexing. In a classified glossary, it is possible to direct users from one record to others in which its entailed terms are defined by means of notations, as illustrated in Annex E. The elegance of defining texts is, admittedly, impaired by this device, but users who want to look up the definitions of entailed terms will be saved the trouble of hunting for them through the index--or by scanning the other records. Another form of textual indexing takes the form of reference, in an index, to material in a document where the term in question has been used. It is especially useful to mark index terms (or page numbers) in such a way that the user will be able to find points in the text where new concepts and terms have been defined. Actually, in the material supplied in Annex F references are given only to the paragraphs in which the index terms are defined--in almost all instances new concepts have been put forward and the terms are stipulated, but in a few cases, relatively unfamiliar words taken from general dictionary entries have been included, with their established definitions.

3.74 Marginal Terms. In a well-designed glossary, it is important to help users find definitions for marginal terms. As in #1.20, marginals are entailed terms that are semantic derivatives, i.e. their definitions will not be found in a general dictionary, but neither will their definitions be found among those of the core concepts of the glossary's subject field, for which defining records are supplied. Two techniques for helping users understand marginal terms are recommended. If their definitions are readily available in a published special dictionary, or in another glossary, an appropriate reference number following each marginal term will guide users to this source. When such a reference tool is not readily available, however, an alternative technique involves the inclusion of definitions for marginal terms in the index. Since the marginal terms are not marked in the networking glossary, this technique cannot be illustrated in Annex F. However, in my report to UNESCO on the INTERCONCEPT project, an index is supplied, which contains for each index term, either a notation (if it is a core term) or a definition (if it is a marginal term).

Conclusion

3.80 No doubt a proper glossary should include elements not covered in this analysis of the networking entries. There should be an introduction, and a list of references. Records should always indicate the source of definitions, and the context of use of the terms offered for each concept. Much more needs to be said about the procedures for working interactively, and continuously, with specialists in the subject field covered by a glossary. The use of computerized terminology banks, and the advantages of computer conferencing, could all be discussed--and they need to be taken into account in any design for an adequate glossary project. Here, however, our attention has been restricted to the physical contents of a glossary, and the advantages, for scholars engaged in the production of knowledge, of a semantic mode that is the inverse of the one that is appropriate for thesaurus design. It is the lack of attention to the theory and practice of glossary design, and the relatively well-developed state of thesaurus design, that has led many glossologists into the selection, by default, of an inappropriate model for their work. If this essay draws attention to the need for and feasibility of an alternative glossary semantic mode, it will have served its purpose.

ANNEX A

PAGES EXCERPTED FROM THE "GLOSSARY FOR LIBRARY NETWORKING"

ISSN 0160-9742

NETWORK PLANNING PAPER

NETWORK DEVELOPMENT OFFICE

NUMBER 2

DATE 1978

TITLE A Glossary for Library Networking

Prepared by Dataflow Systems Inc.

for the

Network Development Office

Library of Congress

Washington

INTRODUCTION

The discussions of the working groups involved in planning the library bibliographic components of a national library network point to the need for standarised vocabulary. Papers presented at working group meetings reflect the same need. The relative newness of the field of library networking and the diversity of background of those working in this area or affected by its development create communication problems. Computer experts are talking to and writing for librarians. Librarians are explaining their needs to telecommunications engineers. Each brings different professional jargon to the combined discipline; each interprets standard terminology according to his own experience, thereby creating a meld of meanings as concepts are combined. Also, other concepts are gaining a new context when viewed from the standpoint of library networking.

The Network Development Office of the Library of Congress responded to this need for commonality of vocabulary by arranging to have this special glossary developed by Dataflow Systems Inc. of Bethesda, Maryland. This is a working document; as usage changes, as new terms evolve to describe different aspects of the field, and as library networking professionals work with the terms on a nationwide scale, the scope of the glossary and the definitions of the terms are expected to change.

An initial glossary was assembled using terms suggested by the Network Development Office and documents in its files, journal articles, minutes of working group meetings, and discussions with individuals. The list was reviewed and revised, and preliminary definitions were developed. The definitions are functional rather than formal. The criteria for inclusion of a term are: that it be in current use relative to library networking, that it be a term unfamiliar to the library community but basic to its comprehension of library networking, or that the term's standard meaning be narrowed or modified by its use in a library networking context.

Preliminary definitions were submitted to the Network Development Office and to designated group representatives for review. Participating in this effort were the Network Advisory Committee, Network Technical Architecture Group, National Commission on Libraries and Information Science (NCLIS), authorities study group, NCLIS/National Bureau of Standards

Task Force on Computer Network Protocols, and Council on Computerized Library Networks. Representatives from each of these groups devoted much time and thought to their review. They offered many constructive comments, from the general and philosophical to specific suggestions regarding style and consistency. After definitions were reworked to reflect their advice, consultations with the group representatives continued with the revised glossary being sent for a second review. Additional highly useful comments were received and, in the main, are reflected in the definitions.

Please note that terms defined in the glossary are underscored when they are used in the body of another definition.

Glossary users may wish to refer to the following sources for definitions used in disciplines other then library networking:

A Guide to Networking Terminology, A. J. Neumann, NBS Technical Note 803, March 1974.

A Compilation of Terms and Definitions Appearing in Z39 Standards, American National Standards (ANSI) Committee Z39, April 1977.

Data Processing Glossary, IBM (GC20-1699-5), October 1977. (Includes definitions from American National Dictionary for Information Processing, Computer and Business Equipment Manufacturers Association, and ISO Vocabulary of Data Processing)

resource sharing (cont'd)

3) sharing of library processing services and the computer facilities, programs, and telecommunications to provide these services.

response time

the elapsed time between an indication of the end of a query and the display of the first character of the response.

ring network

a computer network in which each computer is connected to two adjacent computers in a circular pattern.

service center

see: bibliographic service center.

session

an association between an originating host and a target host through a sequence of control messages that explicitly begin and end the session. A session may have a number of transactions which, in turn, may comprise many messages.

sink

1) a computer or terminal that receives and processes data from a connected channel.

2) the point of usage of data in a network. See also: source.

site data

data sufficient to identify the actual physical location and shelving position of an item. For example, site data may specify special collection sites in single-branch collections and branch sites in multiple-branch collections, and may include call number data.

source

1) a computer or terminal which enters data into a connected channel.

2) the point of entry of data into a network. See also: sink.

specialized common carrier

1) a company authorized by a government agency to provide limited telecommunications services. Examples of specialized common carriers are the value-added networks.

2) the term is also used to connote all those common carriers not covered in the original federal communications legislation.

star network

a computer network in which each peripheral network node is connected only to the computer(s) at a single central facility. See also: centralized (computer) network.

state network

see: intrastate network.

station

an input or output point in a telecommunications system.

target host

1) the host computer in a network session with which the originating host communicates.

2) the destination of a message transmitted in a network from an originating host.

telecommunications

transmission and reception of data by electromagnetic means.

teleprocessing

automated data processing that utilizes telecommunications facilities for data transmission.

TELPAK

commercial telecommunications services using wide-band transmission techniques for multiple channel, high-speed and video communication.

terminal

1) a device for entering data into or receiving data from a computer system or computer network.

2) a point in a communications network at which data can either enter or leave.

text

1) the part of a message that contains the substantive information; the body of a message, exclusive of control characters. See also: message text.

text (cont'd)

2) the accepted format for a standardized network communication, e.g., query text, reply text, error text, etc. See also: message text format.

3) a sequence of characters forming part of a transmission from a data source to a data sink, preceded by a header and followed by an "end of text" signal.

transaction

1) an operational unit of processing at the application level; a complete step of data processing.

2) a logical grouping of messages in both directions between originating hosts and target hosts in a network.

two-way alternate operation

mode of operation of a data link in which data may be transmitted in both directions, one way at a time, over a single channel.

two-way simultaneous operation

mode of operation of a data link in which data may be transmitted simultaneously in both directions over two channels.

union catalog

1) a catalog that describes the contents of physically separate library collections, indicating by means of locations data the libraries in which a given item may be found.

2) subsets of the national library network data base.

a) See: national library network union catalog.

ANNEX B

TRACKED ARRANGEMENT: Selected Records taken from "A Glossary for Library Networking," Washington, D.C.: Library of Congress, 1978. Network Planning Paper.

		Tracking Terms
STEP 1: 'sink' and 'source'		
1a sink	1) a computer or terminal that receives and processes data from a connected channel	terminal -- 2d channel -- 2a (data) -- 2b (network) -- 2c
	2) the point of usage of data in a network. See also: source	
1b source	1) a computer or terminal which enters data into a connected channel	
	2) the point of entry of data into a connected channel. See also: sink	
STEP 2: 'channel,' 'data,' ''network,' and 'terminal'		
2a channel	1) the connection between a message source and a message sink in a communications system	message -- 3c
	2) a means of one-way transmission of data	
2b data	a representation to which meaning can be assigned	
2c network	1) two or more organizations engaged in a common pattern of information exchange through telecommunications links, for some common objectives	links -- 3b telecommunications -- 3f
	2) an interconnected or interrelated group of nodes	(nodes) -- 3e see also: network node -- 3d
2d terminal	1) a device for entering data into or receiving data from a computer system or computer network	
	2) a point in a communications network at which data can either enter or leave	communications-network -- 3a
STEP 3: 'communications network,' 'links,' 'message,' 'nodes' 'telecommunications'		
3a communications-network	the physical means for a group of nodes to intercommunicate data	
3b link	a communications path between two nodes or points	
3c message	1) a single transmission in one direction, consisting of a header and data	header -- 4b
	2) a unit of information transmitted from one node to another on a network	
3d network node	a station, terminal, or communications-computer in a computer network	station -- 4c communications-computer -- 4a
3e node	see: network node	
3f telecommunications	transmission and reception of data by electromagnetic means	

STEP 4: 'communications computer,' 'header,' and 'station'

4a communications-computer	a special purpose computer used to control or format data transmitted between network-nodes	
4b header	the control data that precede a message-text: for example, source or destination code, priority, message type	message text -- 5a text -- 5b
4c station	an input or output point in a telecommunications system	

STEP 5: 'message text,' and 'text'

5a message text	the substantive portion of a network transmission, i.e. blocks of data exclusive of control characters. See also: header	control characters -- 6a
5b text	1) the part of a message that contains the substantive information; the body of a message, exclusive of control characters. See also: message text 2) the accepted format for a standardized network communication, e.g. query text, reply text, error text, etc. see also: message text format	message text format -- 6b

STEP 6: 'control character,' and 'message text format'

6a control-character	a character whose occurrence initiates, modifies, or terminates a control function	control function -- 7a
6b message text-format	standardized format for messages in a network. Types of message text formats include query text, bibliographic text, reply text, error text. A message text format for a network could be: 1) header length, 2) header, 3) data length, 4) data	

STEP 7: 'control function'

7a control function	an operation that affects the recording, processing, transmission, or interpretation of data

CHAINED ARRANGEMENT: The following data, taken from Annex B, have been rearranged in such a way that the terms used in each record are either "primitives" or have been defined in a previous record. The several steps are organized so that they include all the definitions that can be understood with a knowledge of ordinary language terms and the previously defined derivative terms.

No.	Defining Text	Terms	Used in:
STEP 1:			
11	a representation to which meaning can be assigned	data	21-23, 51, 53-56, 61-3, 71, 75, 77, 81,82
12	a communications path between two points	link	41
STEP 2:			
21	a means of one-way transmission of data	channel(2)	61, 62
22	transmission and reception of data by electromagnetic means	telecommunications	31, 41
23	an operation that affects the recording, processing, transmission, or interpretation of data	control function	51
STEP 3:			
31	an input or outpoint point in a telecommunications system	station	64
32	a character whose occurrence initiates, modifies, or terminates a control function	control character	51
STEP 4:			
41	two or more organizations engaged in a common pattern of information exchange through telecommunications links	network(1)	51-55, 64, 72 74, 77, 81
STEP 5:			
51	the substantive portion of a network transmission, i.e. blocks of data exclusive of control characters	message text	56
52	the accepted format for a standardized network communication, e.g. query text, reply text, error text, etc.	text(2)	--
53	the point of usage of data in a network	sink(2)	75, 83
54	the point of entry of data into a network	source(2)	75, 83
55	a device for entering data into or receiving data from a computer system or computer network	terminal(1)	61, 62, 64
56	the control data that precede a message text, for example, source or destination code, priority, message type	header	63, 75, 81
STEP 6:			
61	a computer or terminal that receives and processes data from a connected channel	sink(1)	75, 83
62	a computer or terminal which enters data into a connected channel	source(1)	75, 83
63	a single transmission in one direction, consisting of a header and data	message(1)	73, 81, 83
64	a station, terminal, or communications computer in a computer network*	node, network node	71, 73, 76, 77

* (64) breaks the chain, since it includes communications computer, which has not yet been defined--see(77). Also, it uses network, but possibly in the sense to be defined, below, in (76) There seems to be no way to escape this circularity.

STEP 7:			
71	the physical means for a group of nodes to intercommunicate data	communications-network	82
72	a unit of information transmitted from one node to another on a network	message(2)	73, 81, 83
73	the part of a message that contains the substantive information	text(1)	--
74	the accepted format for a standardized network communication	text(2)	--
75	a sequence of characters forming part of a transmission data source to a data sink, preceded by a header, and followed by an "end of text" signal	text(3)	--
76	an interconnected or interrelated group of nodes *	network(2)	51-5, 64, 72, 74, 77, 81
77	a special purpose computer used to control or format data transmitted between network nodes	communications-computer	64
STEP 8:			
81	standardized format for messages in a network: it could include 1) header length; 2) header, 3) data length, 4) data	message text format	--
82	a point in a communications network at which data can either enter or leave *	terminal(2)	61, 62, 64
83	the connection between a message source and a message sink in a communications system *	channel(1)	61, 62

* Possible circularities are again involved since it is not clear which of the senses of 'network,' 'terminal,' and 'channel' were intended in the prior definitions in which these terms appear. In a perfectly chained sequence, the notations for records in which a term is used should always be higher than the number of the record in which that term is defined.

A TRANSITIONAL CLASSIFIED ARRANGEMENT: All material in these records are quoted, without revision, from the networking glossary

No.	Defining Text	STIC	STOC *
100	ARCHITECTURE (STRUCTURE) OF LIBRARY NETWORKING		
110	Whole Network		
111	"an interconnected or interrelated group of nodes"	network	---
112	"the physical means for a group of nodes to intercommunicate data"		communications-network
120	Network Links		
121	"a communications path between two nodes or points"		link
122	"the assemblage of communications equipment and interconnecting circuits that allows data to be exchanged between two or more stations"		data link
123	"the connection between a message source and a message sink in a communications system"	channel	---
130	Network Nodes		
131	"a station, terminal, or communications-computer in a computer network"		node, or network node
132	"a special purpose computer used to control or format data transmitted between network nodes"		communications-computer
133	"a device for entering data into or receiving data from a computer system or computer network"	terminal	
134	"an input or output point in a telecommunications system"		station
135	"a computer or terminal which enters data into a connected channel"	source	---
136	"a computer or terminal that receives and processes data from a connected channel"	sink	---
200	CONTENTS OF NETWORK TRANSMISSIONS		
210	Transmissions		
211	"a representation to which meaning can be attached		data
220	Format of Transmissions		
221	"accepted format for a standardized network communication"	text	---
222	"standardized format for messages in a network"		message text format

*To overcome ambiguity in the production of knowledge, there should be at least one univocal (STOC) term for each concept in a subject field, like library networking. The dashes indicate that no univocal term is provided by the glossary for the concepts concerned. All terms listed under "STIC" are multivocal--they designate more than one concept in the field of library networking, according to the glossary. It is, of course, unnecessary to have multivocal terms--but they are often convenient to use.

		STIC	STOC
230	Data Elements in Transmissions		
231	"the substantive portion of a network transmission, i.e. blocks of data exclusive of control characters"		message text
232	"a single transmission in one direction, consisting of a header and data"	message	---
233	"the part of a message that contains the substantive information; the body of a message, exclusive of control characters"	text	---
234	"the control data that precede a message text"		header
235	"a sequence of characters forming part of a transmission from a data source to a data sink, preceded by a header and followed by an "end of text" signal"	text	---
240	Non-Data Elements in Transmissions		
241	"a character whose occurrence initiaties, modifies, or terminates a control function"		control character
300	FUNCTIONS (OPERATIONS, ACTIVITIES) OF NETWORKING		
310	User Functions		
311	"two or more organizations engaged in a common pattern of information exchange through telecommunications links for some common objective"	network	---
312	"a means of one-way transmission of data"	channel	---
313	"a point in a communications network at which data can either enter or leave"	terminal	---
314	"the point of entry of data into a network"	source	---
315	"the point of usage of data in a network"	sink	---
320	Transmission Functions		
321	"transmission and reception of data by electromagnetic means		telecommunications
322	"a unit of information transmitted from one node to another"	message	---
323	"an operation that affects the recording, processing, transmission, or interpretation of data"		control function

REVISED CLASSIFIED ARRANGEMENT: Some of this material is quoted from the networking glossary, as shown by quotation marks around the defining texts--revised definitions are given in parentheses. New (univocal) terms are marked with parentheses--other terms are those given in the networking glossary. Numbers in parentheses index the records for defined terms.

		STIC	STOC
100	ARCHITECTURE (STRUCTURE) OF LIBRARY NETWORKING		
110	Whole Network		
111	"an interconnected or interrelated group of nodes (131)"	network	(network structure) (network(2))
112	(a network structure (111) used to communicate data (211))		communications-network
120	Network Links		
121	"a communication path between two nodes (131)"		link, or data link
122	(omitted)		
123	(a link (121) between source (134) and sink-stations (135) in a network structure (111)	channel	(channel structure) o (channel(1)
130	Network Nodes		
131	(a station (134), terminal structure (133), or communications-computer (132) in a communications network (112))		node, or network node
132	(a computer used to control or format data (211) transmitted between nodes (131))		communications-computer
133	(a device for entering data (211) into or receiving data (211) from a network structure (111))	terminal	(terminal structure) (terminal(1))
134	(a terminal structure (133) in a telecommunications (321) system)		station
135	(a station (134) which enters data (211) into a communications network (112))	source	(source station) (source(1))
136	(a station (134) which receives and processes data (211) from a channel structure (123))	sink	(sink station) (sink(1))
200	CONTENTS OF NETWORK TRANSMISSIONS		
210	Transmissions		
211	(a representation to which meaning can be assigned by persons receiving communications)		data
212	(characters used in transmission which are not meaningful to persons receiving communications)		(non-data)
213	(a complete transmission, including data (211) and non-data (212) elements)		(whole text)
220	Format of Transmissions		
221	(accepted format for a standardized communication (322) in a network structure (111))		(whole text format) (text(2) ?)
222	(accepted format for data (211) elements contained in a whole text (213)	text	(message text format) (text(2) ?)

		STIC	STOC
230	Data Elements in Transmissions		
231	(data (211) elements contained in a whole text (213)		message text text(1) ??? message(1) ???
232	(omitted)		
233	(omitted)		
234	(the control data (211) that introduces a message-text (232))		header
235	(the data (211) contained in a message text (231) following the header (234)	text	(running text) (text(3))
240	Non-Data Elements in Transmissions		
241	"a character whose occurrence initiates, modifies, or terminates a control function (323)"		control character
300	FUNCTIONS OF NETWORKING		
310	User Functions		
311	(organizations using a communications network(112)	network	(functioning network) (network(1))
312	(a link (121) used to transmit data (211) one way in a functioning network (311))	channel	(functioning channel) (channel(2))
313	(a node (131) where data (211) can enter or leave a functioning network (311))	terminal	(functioning terminal (terminal (2))
314	(a functioning terminal (313) used to enter data (211) into a functioning network (311)	source	(functioning source) (source(2))
315	(a functioning terminal (313) used by data (211) receivers)	sink	(functioning sink) (sink(2))
320	Transmission Functions		
321	"transmission and reception of data (211) by electromagnetic means"		telecommunications
322	(a unit of information transmitted between nodes (131))	message	(communication) (message(2))
323	"an operation that affects the recording, processing, and transmission or interpretation of data (211)"		control function

*Underlined terms in the defining texts are all univocal (STOC) terms whose definitions may be found at the notation given in parentheses, following the term. If these terms are found in the networking glossary, they are also underlined in the STOC column, but if they were introduced here, for illustrative purposes, they are given in parentheses, without underlining, in the STOC column. Terms given in the STIC column are not used in the defining texts of Annex E. These are the terms given in the networking glossary that are prone to ambiguity unless the context shows very clearly which of their possible meanings, in library networking, is intended.

INDEX: Terms Defined in Annexes B, C, D, and E

TERMS	ANNEX B	ANNEX C	ANNEXES D & E
channel; 1) channel structure	2a	83	123
2) functioning channel	2a	21	312
channel, functioning; channel(2)	---	---	312
channel structure; channel(1)	---	---	123
character, control	6a	32	241
communication; message(2)	---	---	322
communications computer	4a	77	132
communications network	3a	71	112
computer, communications	4a	77	132
control character	6a	32	241
control function	7a	23	323
data; link data	2b	23	323
data, non-	---	---	212(E only)
data link; data	---	---	122(D); 121(E)
format, message text; text(2) ?	6a	81	222
format, whole text ; text(2) ?	---	---	221
function, control	---	23	323
functioning channel; channel(2)	---	---	312
functioning network; network(2)	---	---	311
functioning sink; sink(1)	---	---	315
functioning source; source(2)	---	---	315
functioning terminal; terminal(2)	---	---	313
header	4b	56	234
link	3b	12	121
link, data; data(?)	---	---	122(D); 121(E)

message; 1) message text	3c	63	232
2) communication	3c	72	322
message text; message(1)	5a	51	232
message text format	6a	81	222
network; 1) functioning network	2c	41	311
2) network structure	2c	76	111
network, communications	3a	71	112
network, functioning; network(1)	---	41	311
network structure; network(2)	---	---	111
network node; node	3d	64	131
node; network node	3e	64	131
node, network; node	3d	---	131
non-data	---	---	212(E only)
sink; 1) sink station	1a	61	136
2) functioning sink	1a	58	315
sink, functioning; sink(2)	---	---	315
sink station; sink(1)	---	---	136
source; 1) source station	1b	62	135
2) functioning source	1b	54	314
source, functioning; source(2)	---	---	314
source station; source(1)	---	---	135
station	4c	31	134
station, sink; sink(1)	---	---	138
station, source; source(1)	---	---	135
structure, channel; channel(1)	---	---	123
structure, network; network(2)	---	---	111
structure, terminal; terminal(1)	---	---	130

terminal; 1)terminal structure	2d	55	133
2) functioning terminal	2d	82	313
terminal, functioning; terminal(2)	---	---	313
terminal structure; terminal(1)	---	---	133
telecommunications	3f	22	321
text; 1)message text	5b	73	231
2) whole text format(?)	5b	74	321
3) running text	5b	75	235
text, message; message(1)?; text(1)?	5a	51	231
text, running; text(3)	---	---	235
text, whole; text (?)	---	---	213(E only)
text format, message; text(2) ?	6a	81	222
text format, whole; text(2) ?	---	---	221
whole text;	---	---	213(E only)
whole text format; text(2) ?	---	---	221

*The index terms that were used as multivocal terms in the networking glossary have been underlined, and the illustrative univocal terms that might be used for each of the senses of these terms have been given after them.

ANNEX G: Index of Technical Terms introduced in "The Design of Glossaries"

Stipulated definitions of the following terms will be found at the numbered paragraphs of Part I, as indicated below

compound term , 0.09
condensation, 1.32
core term, 1.20

defined term, 0.06
defining text, 2.27
definition, independent, 1.10
definition, interdependent, 1.13
derivation, 1.12
derivative term, 1.12

elasticity, 1.37
entailed term, 0.05
eolog, 2.48
eonym, 2.48

familiar term, 2.26

glossary semantic mode, 2.34
glossology, 2.04
glossonym, 2.27

higher level derivative term, 1.17

independent definition, 1.10
inelastic, 1.38
interdependent definition, 1.13

lexism, 2.43

marginal term, 1.20
multinomial, 2.23
multisemantic, 2.32
multivocal term, 2.14
multivocalism, 2.12

neolog, 2.48
neonym, 2.48
neoterism, 2.43

primary derivative term,
primitive term, 0.07

reduction, 1.34

secondary derivative term, 1.16
sleeper, 1.25
stretching, 1.32
STIC, 2.28
STOC, 2.28

text, defining, 2.27
term, compound, 0.09
term, core, 1.20
term, defined, 0.06
term, derivative, 1.12
term, entailed, 0.05
term, familiar, 2.26
term, higher level derivative, 1.17
term, marginal, 1.20
term, multivocal, 2.14
term, primary derivative, 1.15
term, primitive, 0.07
term, secondary derivative, 1.16
term, unfamiliar, 2.26
term, univocal, 2.14
term family, 2.25
term replacement, 1.31
thesaurus semantic mode, 2.33
tracking, 1.14

unfamiliar term, 2.26
uninomial, 2.23
unisemantic, 2.32
univocal term, 2.14
univocalism, 2.13